Dissecting the genetic bases of severe malaria resistance using genome-wide and post genome-wide study approaches

By:

Delesa Damena Mulisa

MLSDEL001

Thesis presented for Degree of DOCTOR OF PHILOSOPHY in Human genetics in the Department of Pathology, Faculty of Health Science, University of Cape Town

Supervisor: Associate Professor Chimusa E. Rugamika (University of Cape Town, SA)

September, 2020

ABSTRACT

P. falciparum malaria remains one of the leading public health problems worldwide. The global tally of malaria in 2018 was estimated at 228 million cases and 405, 000 deaths worldwide. African countries disproportionately carry the global burden of malaria accounting for 93% and 94% of cases and deaths, respectively. Even though most infected children recover from *P. falciparum* malaria, a small subset (~1%) of cases progresses to severe disease and death. Over the last decade, several genome-wide association studies (GWASs) have been conducted in diverse malaria endemic populations to understand the natural host protective immunity against severe malaria that can provide clues for the development of new vaccines and therapeutics. However, beyond identifying association variants, conventional GWAS approaches can't inform the underpinning biological functions. To bridge this gap, we applied various contemporary statistical genetic analytic approaches to malaria GWAS datasets of diverse malaria endemic populations.

First, we accessed malaria resistance GWAS datasets of three African populations (N=~11,000) including Kenya, Gambia and Malawi from European Genome Phenome Archive (EGA) through MalariaGEN consortium standard data accession procedures. We explored the challenges of GWAS approaches in the genetically diverse Africa populations and figured out how various advanced statistical genetic methods can be implemented to address these challenges. We investigated single nucleotide polymorphism (SNP) heritability (h^2_g) of malaria resistance in the Gambian populations and determined appropriate quality (QC) thresholds to accurately estimate the h^2_g in our dataset. Second, we estimated h^2_g in the three populations and partitioned the h^2_g into chromosomes, allele frequencies and annotations using the genetic relationship-matrix restricted maximum likelihood approaches. We further created African specific reference panel from African population datasets obtained from 1000 Genomes Project and African Genome Variation Project dataset and computed linkage disequilibrium (LD). We used LD information obtained from these reference panels to compute cell-type specific and none cell-type specific enrichments for GWAS-summary statistics meta-analyzed across the three populations. Our results showed for the first time that malaria resistance is polygenic trait with h^2_g of ~20% and that the causal variants are overrepresented around protein coding regions of the genome. We further showed that the h^2_g is disproportionately concentrated on three chromosomes (chr 5, 11 and 20), suggesting cost-effectiveness of targeting these chromosomes in future malaria genomic sequencing studies.

Third, we systematically predicted plausible candidate genes and pathways from functional analysis of severe malaria resistance GWAS summary statistics (N = 17,000) meta-analyzed across eleven populations in malaria endemic regions in Africa, Asia and Oceania. We applied positional mapping, expression quantitative trait locus (eQTL), chromatin interaction mapping and gene-based association analyses to identify candidate severe malaria resistance genes. We performed network and pathway analyses to investigate their shared biological functions. We further applied rare variant analysis to raw GWAS datasets of three malaria endemic populations including Kenya, Malawi and Gambia and performed various population genetic structures of the identified genes in the three endemic populations and 20 world-wide ethnics. Our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Furthermore, our population genetic analysis revealed that the minor allele frequencies (MAF) of the SNPs residing in the identified genes are generally higher in the three malaria endemic populations compared to global populations. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes that are enriched in pathways linked to severe malaria pathogenesis. This highlights the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways.

In conclusions, this project showed that malaria resistance trait is mainly a polygenic trait which is influenced by genes and pathways linked to blood stage lifecycle of *P. falciparum*. These findings constitute the foundations for future experimental studies that can potentially lead to translational medicine including development of new vaccines and therapeutics. However, '-omics' studies including those implemented in this study, are limited to single data-type analysis and lack adequate power to explain the complexity of molecular processes and usually lead to identification of correlations than causations. Thus, beyond singe locus analysis, the future direction of malaria resistance requires a paradigm shift from single-omics to multi-stage and multi-dimensional integrative multi-omics studies that combines multiple data types from the human host, the parasite, and the environment. The current biotechnological and statistical advances may eventually lead to the feasibility of systems biology studies and revolutionize malaria research.

DECLARATION ON THE THESIS WORK

I declare that this thesis is my *bona fide* work both in concept and execution and that all sources of material used for this thesis have been duly acknowledged. This thesis has been submitted in fulfilment of the requirements for PhD degree at the University of Cape Town. I solemnly declare that this thesis is not submitted to any other institution anywhere for the award of any academic degree, diploma, or certificate.

Signature: Signed by candidate

Date: 07/09/2020

Student Name: ____DELESA DAMENA MULISA

Student Number. MLSDEL001

DECLARATION ON THE INCLUSION OF PUBLICATIONS

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publications in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publications:

1. **Delesa Damena**, Francis Agamah, Peter O. Kimathi, Ntumba E. Kabongo, Hundaol Girma, Wonderful T. Choga, Lemu Golassa and Emile R. Chimusa. Functional analysis of genome-wide dataset from 17000 individuals identifies multiple candidate malaria resistance genes enriched in malaria pathogenic pathways. Preprint: medRxiv preprint doi: https://doi.org/10.1101/2020.08.15.20175471.
2. **Delesa Damena** and Emile Chimusa Genome-wide heritability analysis of severe malaria susceptibility and resistance reveals evidence of polygenic inheritance Human Molecular Genetics, 2020, 29:168–176, doi: 10.1093/hmg/ddz258
3. **Delesa Damena**, Denis Awany, Lemu Golassa and Emile Chimusa. Genome-Wide Association Studies of malaria susceptibility and resistance: progress, pitfalls and prospects.BMC Medical Genomics (2019) 12:120 https://doi.org/10.1186/s12920-019-0564-x

Signature: Signed by candidate

Date: 07/09/2020

Student Name: DELESA DAMENA MULISA

Student Number: MLSDEL001

Publication from collaborative efforts during PhD project

4. Francis E. Agamah, **Delesa Damena**, Michelle Skelton, Anita Ghansah, Gaston K. Mazandu, Emile R. Chimusa. Large–scale data–driven network analysis of human-plasmodium falciparum interactome: processes for malaria drug discovery and extracting in silico targets. Accepted by Hum Genet. journal
5. Emile R. Chimusa, Joel Defo, Jacquiline W. Mugo, Denis Awany, **Delesa Damena**, Imane Allali Ephifania Geza, Samar M. Elsheikh, Hassan Ghaza, Ahmed Moussa and Gaston K. Mazandu. Dating admixture events is unsolved problem in multi-way admixed populations. Briefings in Bioinformatics, 00(00), 2018, 1-12.
6. **Delesa Damena**, Samuel Tolosa, Milkessa Hailemariam, Aboma Zewude, Adane Worku, Biruk Mekonnen, Temesgen Mohammed, Addisu Admasu, Adane Mihret, Emile R. Chimusa, Tamrat Abebe, Gobena Ameni. Genetic diversity and drug resistance characteristic of Mycobacterium tuberculosis isolated from pulmonary tuberculosis patients at Saint Peter TB specialized Hospital, Ethiopia, PLoS ONE 14(6): e0218545. https://doi.org/10.1371/journal.pone.0218545

DEDICATION

In loving memory of my mother Bezunesh Henjigu and my sister Wesene Damena. Through the hardships you endured and the sacrifices you made, I reached this level. I promise that I will continue working hard to make you proud in heaven.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor Prof. Chimusa E. Rugamika for his guidance and immense support throughout the course of this research work. I thank all the staffs in Kwiatkowski's group at Big Data Institute (BDI) during my stay at the University of Oxford. Specifically, I would like to express my heartfelt thanks and appreciation to Dr Gavin Band for helping me develop big genomic data analysis skills. My sincere gratitude goes to all the staffs and students of the division of Human genetics for their immense support and encouragements throughout my stay in the laboratory. My outpouring love and respect go to my father Damena Mulisa and sisters: Gudo Damena, Fasika Damena and Meseret Damena and Sanait Mulisa who are sole sources of my happiness and achievements. I respectfully thank the MalariaGen scientists who generated the original dataset used in this study and the study participants without whom this work couldn't have been accomplished.

Finally, I would like to thank The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) program for providing funding, trainings and related assistances during the course my studies. I also thank Newton's international student transfer scheme for funding me during my stay at University of Oxford.

TABLE OF CONTENTS **PAGE**

LIST OF FIGURES **PAGE**

LIST OF TABLES **PAGE**

In CHAPTER TWO

In CHAPTER THREE

In CHAPTER FOUR

LIST OF ABBREVIATIONS

AGVP …… .African Genome Variation Project

BC……..……Before Christ

CD…….……Cluster of Differentiation

CIDR……Cysteine-Rich Inter Domain Region

CM…..……..……..…….Cerebral Malaria

CR1…..……Complement Recepter1

DBL ……..……Duffy-Binding Like Domain

EBP……..…..……..Erythrocyte Binding Proteins

ENCODE…….Encyclopaedia of DNA Elements

eQTL…….. Expression Quantitative Trait Locus

EVs…….……..….Extracellular Vesicles

FUMA…….……Functional Mapping and Annotation of Genetic Association

G6PD…….……..……. Glucose-6-Phosphate

Hb……. Haemoglobin

Hp…….…….. Haptoglobin

h2g……..……..……SNP_heritability

ICAM 1…..…….Intracellular Adhesion Molecule 1

RBC…….……Red Blood Cells

IL…….…….Interleukin

.

GCTA……..……Genome-Wide Complex Trait Analysis

GO……..Gene Ontology

GRM…….……...Genetic relationship matrix

GTEx…….……Genotype-Tissue Expression

GWAS……. ...Genome-Wide Association Studies

LD…… Linkage Disequilibrium

LM…….…… Lymphotoxins

LMM…….…… Linear Mixed Models

MAF……...Minor Allele Frequency

MalariaGEN …….….Malaria Genomic Epidemiology Network

GeneMANIA……Multiple Association Network Integration Algorithm

PASCAL…….……....……Pathway Scoring Algorithm

PCA……. Principal Component Analysis

PfEMP……. Plasmodium falciparum Erythrocyte Membrane Protein

PVM……..……..Parasitifourous Vacuole Membrane

RBC…….……Red blood cells

RELM……..……. Restricted Maximum Likelihood

SKAT. …….Sequence Kernel Association Test

SMA ...……...Severe Malarial Anaemia

SNP…….Single Nucleotide Polymorphism

TF……..……Tissue Factor

TNF ……..Tumour Necrosis Factor

VEGFC ……..……...Vascular Endothelial Growth Factor C

VEGFD…….…….Vascular Endothelial Growth factor D

vWF ……..…….von Willebrand Factor

WHO…….…….....World Health Organization

WPBs …….…….……Weibel–Palade Bodies

1. CHAPTER ONE: GENERAL INTRODUCTION

"God be praised. At His command
Seeking His secret deeds
With tears and toiling breath
I find thy cunning seeds
O million-murdering death"
(Ronal Ross, 1902), Nobel prize winner in Medicine for the discovery of malaria parasite

1.1. A glance in history of malaria, "king of diseases"

Malaria is caused by infection with protozoan parasites belonging to the genus *Plasmodium* transmitted by female Anopheles. It is one the most important diseases of mankind which shaped the course of human history for centuries. Malaria has been documented as an epidemic with symptoms of typical fever and enlarged spleen in early civilized societies including Greek, Roman, Indian, Arabs, Egyptian and Chinese [1]. In old Europe, different forms of malarial disease were given different names including "benign tertian" (*P. vivax*), "quartan" (*P. malariae*), "subtertian, malignant" periodic fevers (*P. falciparum*). "tertian and "quartan" refers to the relapsing fever episodes in every three and every four days, respectively. Similarly, "benign" and "malignant" refers to mild disease and severe life threating disease manifestations, respectively [2]. In the ancient Indian scriptures, malaria was named as "king of diseases" referring to its epidemic nature and deadliest outcome [2].

Malaria is believed to be emerged out of Africa, in the Ethiopian region and spread to the warmer parts of the rest of the world in the past five million years [2,3]. From the valley of Nile, it spread continuously with human mobility to Mediterranean region, to Asia and to the north Europe assisted by agricultural developments [3]. With its global spread, malaria constituted one of the greatest challenges of early civilizations and often responsible for decline of kings and nations [2,3]. Even though controversial, Alexander the Great is said to have died of malaria [2,3]. One of the greatest King of Rome, Gregory the Great (c 540-604), who suffered from malaria described the epidemic fever occurred in Rome in 599 AD*: "For every day I am weak and in pain and sigh, waiting for the remedy of death, assuredly among the clergy of this city and people there are so many cases of lethargy and fever that hardly single*

man remains, who is fit for any office or ministry. However, from neighbouring towns I receive reports every day of the carnage of death" [1].

Malaria was mainly conceived among early populations as the revelation of the supernatural demons that disseminate fevers to bring down kings, military lords and popes [4]. Thus, there had been enormous attempts to defeat malaria through supernatural practices and religious rites by different populations of the old world. The first scientific approach and detailed description of malaria was made by Hippocrates, the greatest scientist of the 5th century, who described malaria a dangerous fume (miasma) that originates from swamps and transported by winds [4].

After periods of several centuries, the discovery of bacteria by Antoni van Leeuwenhoek in 1676 and the emergence of germ theory disproved the belief that malaria is caused by miasma rising from swamps [1]. In 1880, Charles Lous Alphonse Laveran discovered the malaria parasites in red blood cells of infected patients. The discovery was first rejected by many eminent microbiologists of the time including Louis Pasteur; but later accepted after proved by other scientists [4]. Levern was awarded the Nobel Prize for Medicine in 1907 for his work on malaria [5]. In 1897, Ronald Ross, a military surgeon in India discovered the fact that mosquitoes are the vectors of malaria parasites [4]. In the same year, three Italian researchers, Giovanni Battista Grassi, Amico Bignami and Giuseppe Bastianelli described the developmental stages of malaria parasites in the mosquito vectors [4].

More than 4 millennia were required until malaria was fully demystified; the humanity struggled to face one of the most debilitating diseases of mankind which prevented any economic progress in vast regions of the earth for centuries [2]. After Second World War, malaria eradication was accomplished in Europe through implementation of mass drug treatment and residual insecticide spraying [6]. In 1960s, all these areas had been formally declared malaria free by the World Health Organization [7]. In the era of the global malaria eradication, there are still far more fights to defeat this scourge of mankind; particularly, in sub-Saharan Africa where the burden is the greatest.

"But all the evidence we possess would seem to indicate not that poverty is responsible for malaria but that malaria maintains poverty" **Paul Russell in1994**

1.2. The current situation: Progress and challenges of malaria elimination

Malaria is still one of the global health problems with 228 million cases and about 405, 000 deaths in 2018 [8]. African countries disproportionately carry the global burden of malaria accounting for 93% and 94% of cases and deaths, respectively [8]. Malaria has been considered as one of the poverty diseases as it affects the vulnerable populations with underdeveloped healthcare system and poor infrastructures. However, it can be argued that the malaria itself drives poverty not the vice versa. *"But all the evidence we possess would seem to indicate not that poverty is responsible for malaria but that malaria maintains poverty" Paul Russell in1946* [6]. Malaria impacts the economic growth of the endemic countries by weakening production and productivities, incurring costs of treatments, impeding trade and investments [9]. Besides, it has several indirect long-term consequences in the affected children [9]. About 20% of children who recover from severe malaria develop neurological sequalae and sub-optimal neuronal development leading to low performance in school and inadequate self-awareness [9]. By controlling for other factors, it was estimated that a ten percent decrease in malaria incidence in an endemic country would correspond to 0.3 percent increase in economic growth [10].

After the failure of the malaria eradication campaign in 1970s; in fact, which didn't include any of the African countries where the malaria burden is the greatest, a new hope of global malaria elimination become a real possibility in recent years [11]. The global malaria eradication program, accelerated by the WHO led Roll Back Malaria (RBM) partnership, is largely focusing on scaling up the coverage of the available intervention strategies [11]. These include distribution of long-lasting insecticide treated nets (LLINS); indoor residual insecticide spraying, intermittent treatment for pregnant women in high transmission settings and rapid diagnosis and effective treatments using artemisinin-based combination therapies (ACTs) [11]. This led to the significant decline of malaria burden in many parts of the endemic regions [12]. For instance, in between 2010 and 2015, the incidence of malaria and mortality rate among population at risk was globally decreased by 21% and 29%, respectively [13]. In the same period, the under 5 deaths due to malaria fell down by 29 % globally [13]. The RBM partnership developed an action plan for 2016-2030 period targeting a global reduction of at

least 90% in malaria case incidence , mortality rates, and elimination in at least 35 countries by 2030 [14].

Despite such an impressive progress towards global malaria elimination, there are wide arrays of barriers that affect the effectiveness and sustainability of the intervention strategies [15,16]. These include decreased funding, lack of political commitments in some countries, emergence of drug resistant parasites, emergence of insecticide resistant mosquitoes and lack of effective vaccine among others [15,16]. The massive use of insecticide led to the development of resistance to pyrethroids, the only class of compounds used in treating bed nets [17] while resistance to ACTs has been spread in Southeast Asia [18]. The greatest concern of the global malaria eradication program is the possible emergence and spread of the resistance parasite strains to sub-Saharan Africa.

To overcome these challenges and maintain the current momentum, there is a pressing need of new intervention tools including effective drugs and vaccines, rapid and accurate diagnostic tools, effective insecticides, improved surveillances and rapid responses among others [19]. Following the recent advances in molecular biology and bioinformatics, genetic and genomic studies of malaria parasites, mosquitoes and human host are increasingly becoming the critical component of malaria research [20,21]. The increased availabilities of big genomic dataset with improved capacity to analyze and interpretation have led to the practical applications of genomic methods in malaria eradications [21]. These include tracking transmission dynamics and genetic diversity of the parasites and mosquitos, detection of drug resistant parasites and vectors, understanding the impact of drug and insecticide selection on the genomes of malaria parasites and vectors among others [21].

1.3. The parasite: An overview of lifecycle

Malaria is caused by infection with protozoan parasites belonging to the genus *Plasmodium* transmitted by female Anopheles [22]. Transmitted by different species of female mosquitoes, the genes *Plasmodium* infects birds, reptiles and mammals [22]. Out of roughly 200 recognized *Plasmodium* species, five are known to infect humans: *Plasmodium falciparum (P.falciparum), P.vivax, P.malariae, P.ovale and P.knowsi*. The life cycle of the genus *Plasmodium* is complex and alternate between vertebrate and invertebrate hosts **(Fig 1).** The humans and other vertebrates act as the intermediate host for the parasite, while the mosquito, in which the sexual

reproduction takes place, is considered to be the definitive host [23]. I first describe the asexual life cycle in human host and extend to the sexual life cycle in mosquitoes.

1.3.1. Asexual Life Cycle

1.3.1.1. Pre-erythrocyte stage

Female anopheles probes its proboscis in to the host's skin and deposit saliva to prevent the blood from coagulation during the blood meal [24]. In this process, infected mosquitoes inoculates the transmissive form the parasite, the sporozoites in to the skin [24]. From the skin of the infected individual, the sporozoites inter in to the blood circulation or up-taken by the lymphatic systems from which they invade liver [24]. The sporozoites are highly mobile and invasive; they penetrate hepatic macrophages (Kupffer cells), transverse the endothelium protection, scape the lysosomal protection and reach the liver parenchyma [25,26]. The invasion of hepatocytes involves surface proteins of sporozoite (thrombospondin domain on the circumsporozoite protein and on thrombospodin-related protein) and a host surface molecule, heparin sulphate proteoglycans [25,27]. Sporozoites develop in to the pre-erythrocytic schizonts which take 4-15 days depending on the *Plasmodium species* [25]. *P.vivax* and *P.ovale*, may not achieve maturation to a schizont immediately but remain dormant (hypnozoite) [28]. Hypnozoite may remain in the liver for weeks to many years before developing to pre-erythrocytic schizont and lead to relapsing of malaria disease [28]. Although *P. falciparum* is not considered to have a hypnozoite form, it may occasionally demonstrate dormancy [29].

In the hepatocytes, the parasite forms parasitophorous vacuole and differentiates in to trophozoites [26]. After maturation, the parasite buds off the hepatocytes in merozomes containing hundreds of thousands of merozoites [28]. The merosomes lodge in the pulmonary capillaries and slowly disintegrate within 48-72 hours releasing merozoites. The erythrocyte invasion is enhanced when blood flow is slow and the cells are tightly packed [28]. Detailed description of pre-erythrocyte stage has been reviewed in [26].

1.3.1.2. Erythrocyte stage

The erythrocyte stage also termed as blood stage parasites undergo three distinct developmental stages: the merozoite, ring, trophozoite and schizont stages [30]. Merozoite is oval shaped, small cell with approximately a length and a width of 1.6 μm and 1.0 μm, respectively and yet possess all special organelles (rhoptries, microneme and dense granules) needed to invade erythrocytes [30]. The invasion process requires release of essential molecular factors from the organelles including adhesive proteins, proteases and membrane-altering agents [31,32]. Erythrocyte invasion is a complex, multistep process that involve initial binding, reorientation and erythrocyte deformation, junction formation and entry [33]. The primary recognition of erythrocytes and adherence can occur at any point on the surface of the merozoite and is of low affinity [30]. The initial interaction primary involves the parasite's merozoite surface protein (MSPs) and host receptor erythrocyte glycoprotein band 3. Other parasite ligands and host receptors are believed to be involved in this process [30].

This is followed by reorientation events in which the apical end of the parasite is placed in contrast with surface of erythrocyte. After reorientation events, a moving tight junction is formed between the apical end of the parasite and surface of erythrocyte [34]. Aided by the tight junction, parasite enters in in to a parasite induced vacuole formed from erythrocyte membrane (PVM) [31]. The merozoites specifically infect erythrocytes within 30 seconds and requires multiple receptor-ligand interactions [27]. Unlike the *P.vivax* that utilizes Duffy antigens for erythrocyte invasions and restricted to young erythrocytes (reticulocytes), *P. falciparum* use alternative invasion pathways to infect erythrocytes of all ages [35]. These ligands belong to a family of protein called erythrocyte binding proteins (*EBP*) including Duffy binding like (*EBA-175*, *EBA-140*, *EBA-181* and *EBL1*) and reticulocyte binding like homologs (*PfRh2a, PfRh2b and PfRh4*) families [36]. The majority of the ligands are functionally redundant which increases the diversity [37]. Multiple erythrocyte receptors are involved in the invasion process including *Glycophorin A, B* and *C, CD36, ICAM* among others as reviewed elsewhere [38]. After successful invasion, the merozoites loses its organelles and develops to 'a ring stage'. The term 'ring' refers to its disk-like structure under light microscope. The ring stage parasite feeds haemoglobin and develops in to trophozoites, a plumper like cell, which veraciously degrade more haemoglobin and uptakes the host cytosol to its food vacuole [39]. The detailed description of the asexual life-cycle is reviewed in [38].

1.3.2. Sexual life cycle

In the human host, a small proportion of merozoites differentiate in to sexual forms (gametocytes) macrogametocyte (female) and microgametocyte (male) [40]. The development and maturation of gametocytes in *P. falciparum* is slower (12 days) compared to other *Plasmodium* species which takes only about 30-55 hours [40]. It has been postulated that stress factors trigger the parasite to switch to the sexual phase of development [41]. The sexual phase development involves five distinct stages (I-V) marked by expressions of various gametocyte-stage specific proteins (*Pfs16, PF14 744, PF14 748, Pfg27*) and more than 600 gametocyte-sex specific proteins [41]. The ultrastructure of each developmental stage has been shown in [42].

Mature gametocytes are large sized parasites which occupy almost the entire host erythrocyte before they egress [43]. Immediately after ingestion by mosquito, gametocytes escape from their surrounding erythrocyte membrane to undergo gametogenesis [43]. The process of gametogenesis is believed to be activated by a drop of temperature and chemicals in the mosquito host [43]. The male gametocyte multiply itself to multiple ex-flagellated gametes and fertilize the female gametes. The fertilized zygote develops to an invasive motile Ookinete that penetrate epithelial cell layer and exit the gut lumen [44]. In the midgut epithelium, the Ookinete transforms a cyst like structure called oocyst in which the parasite undergoes sporogonic replications and eventually transformed to the infective sporozoites and migrate to salivary glands. The detailed sexual life cycle in mosquito is reviewed in [45].

1.4. Pathogenesis of *P. falciparum* severe malaria

1.4.1. Severe malaria: clinical syndrome

Malaria is a complex disease with clinical manifestations ranging from asymptomatic to life-threating severe cases [46]. In endemic populations, only a small fraction (1-2%) of cases progress to severe malaria (SM) [47]. Nevertheless, severe malaria is one of the commonest reasons for admission to hospital and is a major cause of hospital death in children aged 1-5 years in endemic areas [48]. Severe malaria is defined as demonstration of asexual forms of the malaria parasites in the blood of a patient with a potentially fatal manifestation or complication of malaria in whom other diagnosis have been excluded [49,50].

The major complications of SM include cerebral malaria (CM), pulmonary oedema, acute renal failure, severe malarial anaemia (SMA). Acidosis and hypoglycaemia are the most common metabolic complications which can rapidly develop and progress to death within few hours or days [49]. About eighty percent (80%) of the deaths associated with malaria are caused by CM [49,51].

CM is defined as coma with the presence of *P. falciparum* in the blood of a patient with no evidences of other cause of coma including hypoglycaemia, meningitis, or a postictal state [49]. In endemic areas with intense transmission, CM is characterized a brief fever and generalized illness which rapidly develops to coma. The coma threshold in CM is defined as Blantyre coma score < 2 in children or Glasgow coma score < 11 in adults [49].

Besides, it has been shown that there is a direct link between CM and malarial retinopathy (retinal whitening, vessel colour changes, white-centred haemorrhages) resulting from the sequestration of the infected RBC in microvasculature of the brain and associated pathologies [52]. Malarial retinopathy is often used as an in vivo surrogate marker of CM [53]. In the presence of the other criteria, malarial retinopathy dramatically improves the specificity of CM diagnosis from approximately 61% to 100% [54]. However, CM cases with viral co-infection may not manifest retinopathy. Brian swelling has also reported as one of the cardinal clinical manifestations of CM [53]. CM cases with brain swelling are likely to progress in to brain dysfunction and respiratory failure which leads to fatal outcome. Children with CM in the absence of brain swelling were shown to have higher chance of survival [53]. About 10% of the survivors of CM develop permanent or transient disabilities [52].

SMA is another complication of severe malaria, particularly in children and pregnant women [49]. SMA is defined as haemoglobin < 5 g/100 ml or haematocrit < 15% and other malaria related syndromes [49]. The causal mechanisms of SMA is attributable to several factors including erythrocyte lysis during the parasite replication, sequestrations, splenic removal of deformed iRBCs, suppression of haematopoiesis, immune mediated haemolysis and RBC eryptosis among others [55]. The destruction of uninfected RBCs constitute about 90% of the total loss of RBCs during SM [56].

The clinical manifestations of CM and SMA anaemia appears to be different in that SM usually either leads to CM or SMA but not both at the same time. It has been suggested that the mismatch is due to the variations at which the iRBCs are removed from the circulation by spleen [57]. Effective splenic clearance of immature iRBCs decreases the rate of sequestration and hence protect against CM while excessive removal of un-infected RBCs may lead to SMA. On the other hand, slow and less active splenic removal of RBCs may facilitate sequestration

and leads to rapid development of CM while it decreases the risk of SMA [57]. However, severe malaria cannot be fully explained by these clinical syndromes. Rather, it is constituted by a combination of multiple distinct pathophysiological processes that lead to complications and reduced oxygen delivery to tissues [37,55].

1.4.2. Pathogenesis of severe malaria: Molecular mechanisms

1.4.2.1. Adhesion of infected erythrocytes to host tissues

Although better insights have been gained in recent years, the molecular bases of SM pathogenesis remain obscure due to the complex nature of the parasite life cycle, host genetic factors and immune responses [37]. However, the pathologies observed during malaria infections are generally the consequence of erythrocytes infections by merozoites also known as blood stage life cycle of the parasite (**Fig 1**). *P. falciparum* infected red blood cells (RBCs) adhere to vascular endothelium [58]. This phenomenon, commonly called "sequestration" is believed to be the mechanism of avoidance from splenic clearance and host immunity [59]. Parasite adhesions phenotypes include binding to endothelial cells (cytoadherence) [60], binding with uninfected erythrocytes (rosetting) [61] and platelet-mediated clumping of the infected cells [62]. All adherence phenotypes are resulted from the structural alteration of the infected erythrocyte surface commonly called 'knobs'[63]. Knob is an electron dense protrusion induced by *P. falciparum* derived array of proteins [63]. The major protein family that orchestrates the knob formation and the subsequent sequestrations is a *P. falciparum* erythrocyte membrane protein1 (*PfEMP1*) [63].

PfEMP1 is encoded by approximately 60 genes collectively known as *var* family [64,65]. The *var* genes are one of the multi-gene families that exhibit extremely variable gene expressions patterns which modify the antigenic and functional properties of infected erythrocytes during the mature stages of the intraerythrocytic cycle [66]. The *var* genes demonstrate mutually exclusive clonal gene expression mechanisms in which one *PfEMP1* protein is expressed by any individual parasite and reserve the remaining repertoire [66]. The extracellular region of *PfEMP1* contain variable adhesion domains including Duffy-binding like (*DBL*) domain and 1-2 cysteine-rich inter domain region (*CIDR*) reviewed in [37].

Studies have showed that there are three distinct *var* gene categories named as A, B and C groups on the basis of the conserved upstream region [67,68]. Among these groups, expression of group 'A' genes have stronger associations with SM [68]. Subsequent studies showed that only a subset of group 'A' genes containing binding domain: *DC8* or *DC13*, at their N-terminal

ends are associated with the severity of malarial disease [69,70]. *PfEMP1* demonstrate variations in binding specificity to different host receptors. *Chondroitin sulphate A* is the main receptor of infected erythrocytes in placenta (pregnancy malaria) [71]. *ICAM1* is the commonest receptor in the brain tissue (cerebral malaria) [72]. *Complement recepter1 (CR1*), *heparin sulphate* and *blood group A and B* receptors are shown to be associated with the rosetting which lead to SMA as reviewed in [58].

There have been several other host receptor molecules identified thus far including *CD36, CD56, P-selectin, PECAM1, Thrombospondin, Fractalkine, gC1qR/HABP1/p3* among others [58]. While non-adherent RBCs are cleared by the spleen rapidly, the iRBCs remain protected in the site of their adhesion [37]. Sequestration disrupts blood flow, damages endothelial cells, promote inflammatory cells activation and may lead to severe disease [58]. However, the adhesion phenotypes are different based on the parasite strains and the involved host tissues and thus, not all adhesion lead to life-threatening disease. Different strains of parasites can bind to a combination of different receptors in different tissues and hence, the severity of the disease is heterogeneous [37]. The parasite may demonstrate different antigens in each asexual life cycle which results in a new adhesion phenotype [73]. Furthermore, the heterogeneity of adhesion phenotypes based on geographic regions and transmission intensity have been reported [58].

1.4.2.2. Inflammatory responses

Another important event during SM is the massive release of pro-inflammatory cytokines including *interleukin (IL)-1, IL-6, tumour necrosis factor (TNF), lymphotoxins (LM), interferon gamma (IFNγ) and superoxide* [74]. The early stages of pro-inflammatory cytokines production during the malaria infection plays protective roles. For instance, the pyrogenic cytokines such as *IL-1β* and *TNF* determines the cyclic paroxysms of malaria parasite. This may help to delay the growth of the parasite by affecting its temperature preferences [75]. However, in the later stages, over reaction of the inflammatory reaction is one of the critical factors that drive malaria pathologies including vascular dysfunctions and organ damages, suppression of erythropoiesis and enhance the expression of adhesion molecules [74].

The release of pro-inflammatory cytokines is thought to be mediated mainly by parasite-driven factors released during erythrocyte invasion events, rupture of iRBCs, rupture of schizonts and sequestration processes [52]. The parasite molecules expressed in iRBCs trigger innate immunity that leads to splenic removal of the iRBCs. However, the splenic removal process

triggers the cytokines production and is the major driving factor of the cytokine storm during the malaria episodes [51]. In addition to this, the erythrocyte invasion events that involve interactions of the parasite glycosylphosphatidylinositol (GPI)-anchored Merozoite Surface Proteins (MSP-1 and -2) and host receptors is one of the inducers of *TNF-α* [76]. *TNF-α* in turn induces the epithelial expression of *ICAM-1* in the brain which facilitates sequestration of iRBCs [76]. The release of *TNF-α* stimulates the induction of *IFN*γR1 from natural killer cells [77]. Polymorphisms of *IFN*γR1 gene have been reported to confer resistance against CM in Gambian populations [78]. It has been shown that *LM* which also share the same receptor, is principal mediator of murine CM instead of *IFN*γ [79].

The asexual reproduction of the parasite (schizogony) remarkably increases histone content in its cell which eventually accumulate in the cite of parasite sequestration and stimulate the production of *IL-8* and other inflammatory mediators [80]. A recent study further implicated that the parasite histone contributes to the thrombosis and brain swelling and associated pathology during CM [81]. Another important parasite-driven protein which induces pro-inflammatory cytokine is a histidine-rich protein 2 (*HRP2*) [82]. *HRP2* is produced by the parasite in the iRBCs [83] and accumulates in the site of sequestration in the brain where it induces inflammasome and causes vascular leakage and associated pathologies [82]. It has been shown that *HRP2* level in the plasma correlates with the progression of CM [84]. Furthermore, the parasite genome which is known to be rich in CpG motifs, is one of the major inducers of innate immunity [85]. It has been shown that the parasite DNA bounded by the by-product of digested haemoglobin (hemozoin) in iRBC triggers the release of a proinflammatory cytokines [86]. Similarly, an invitro study has shown that colocalized DNA with hemozoin induce other systemic inflammation such as *NLRP3* and *AIM2* inflammasomes during malaria infection [87].

In addition to the parasite factors, the host-driven components contribute to the induction of pro-inflammatory cytokines and thereby determine the disease severity. A classic example of the host-driven factor is the release of extracellular vesicles (EVs) from iRBCs. EVs, also termed as micro vesicles, are small vesicles containing various molecules including proteins, RNA, and even organelles [88]. The EVs are involved in various biological activities such as transferring genetic information between cells [88]. Studies have shown that EVs are able to activate innate immunity both in mice and humans and are linked to the malaria disease severity [89,90]. It has been further showed that EVs are actively involved in signalling and

cellular communications within the parasite populations and externally with the host immune system [89,90]. A recent study implicated that iRBC-driven EVs contain several important molecules including parasite proteins, DNAs and regulatory microRNAs [91] and form a functional RISC complex with human protein called Ago2. This complex shown to have the capacity of silencing endothelial gene expressions and leads to vascular disfunctions [91]. Likewise, Platelet-driven EVs can effectively link with iRBCs and various endothelial cells and play critical role in malaria pathogenesis [92].

Another important host-driven factor that is capable of eliciting inflammatory reaction is a free heme released to extracellular environment after parasite-induced haemolysis [93]. The oxidation of free haemoglobin leads to the production of reactive oxygen species including *superoxide anions (O_2), hydroxyl radicals (OH), and hydrogen peroxide (H_2O_2)* [39]. These degradation products mediate oxidative stress and induce inflammations resulting in tissue damage, increase expression of adhesion molecules, vascular leakages and infiltration of leukocytes [93]. However, the toxic effects are shown to neutralized by the action of *heme-oxygenase 1* enzyme which degrades heme in to none toxic products including iron (Fe), biliverdin and carbon monoxide (CO*)* [94].

1.4.2.3. Dysregulation of coagulation cascades

In addition to sequestration of iRBCs and inflammatory responses, dysregulation of coagulation cascades is one of the key factors that drive the pathogenesis of SM. Several studies have reported the presence of thrombi and haemorrhages in different tissues of infected children including in brain, kidney, lungs and retina [95,96]. The activation of coagulation cascade is thought to be resulted from malaria-related pathological events such as the release of pro-inflammatory cytokines, endothelial damages, activation of platelets and binding of coagulation-based receptor with iRBCs [97]. Activated endothelium during CM has been shown to induce elevated secretion of von Willebrand factor (vWF), a large protein derived from endothelial surface and megakaryocytes and stored in Weibel–Palade bodies (WPBs) [98]. The abnormal accumulation of vWF induces platelet-mediated clumping of iRBCs in the brain microvasculature [98,99].

In addition to vWF, elevated secretions of tissue factor (TF) has been shown to play a critical role in initiating blood coagulation during SM [100]. The secretion of TF is enhanced by the release of TNF and decreased vascular bioavailability of nitric oxide (NO) during malaria

infections [97]. The free Hb released from parasite-induced haemolysis during SM quenches the NO from the circulation [101]. The resulting decrease in endothelial bioavailability of NO leads to impaired neuronal signalling, oxidant damage and vascular dysfunction [102,103]. Platelets are another important factor which promote blood coagulations during SM [92]. Activated by the inflammatory cytokines, platelets accummulate of at the site of adhesion and promote blood coagulation activities [104,105]. It has been shown that interactions between the platelets and endothelial cells play vital role in disruptions of blood-brain barrier (BBB) and associated complications during CM [106]. Taken together, the molecular mechanisms of SM pathogenesis involve combinations of sequestrations, inflammatory responses and dysregulation of haemostasis. These molecular events are not mutually exclusive and are driven by various parasite factors and host components.

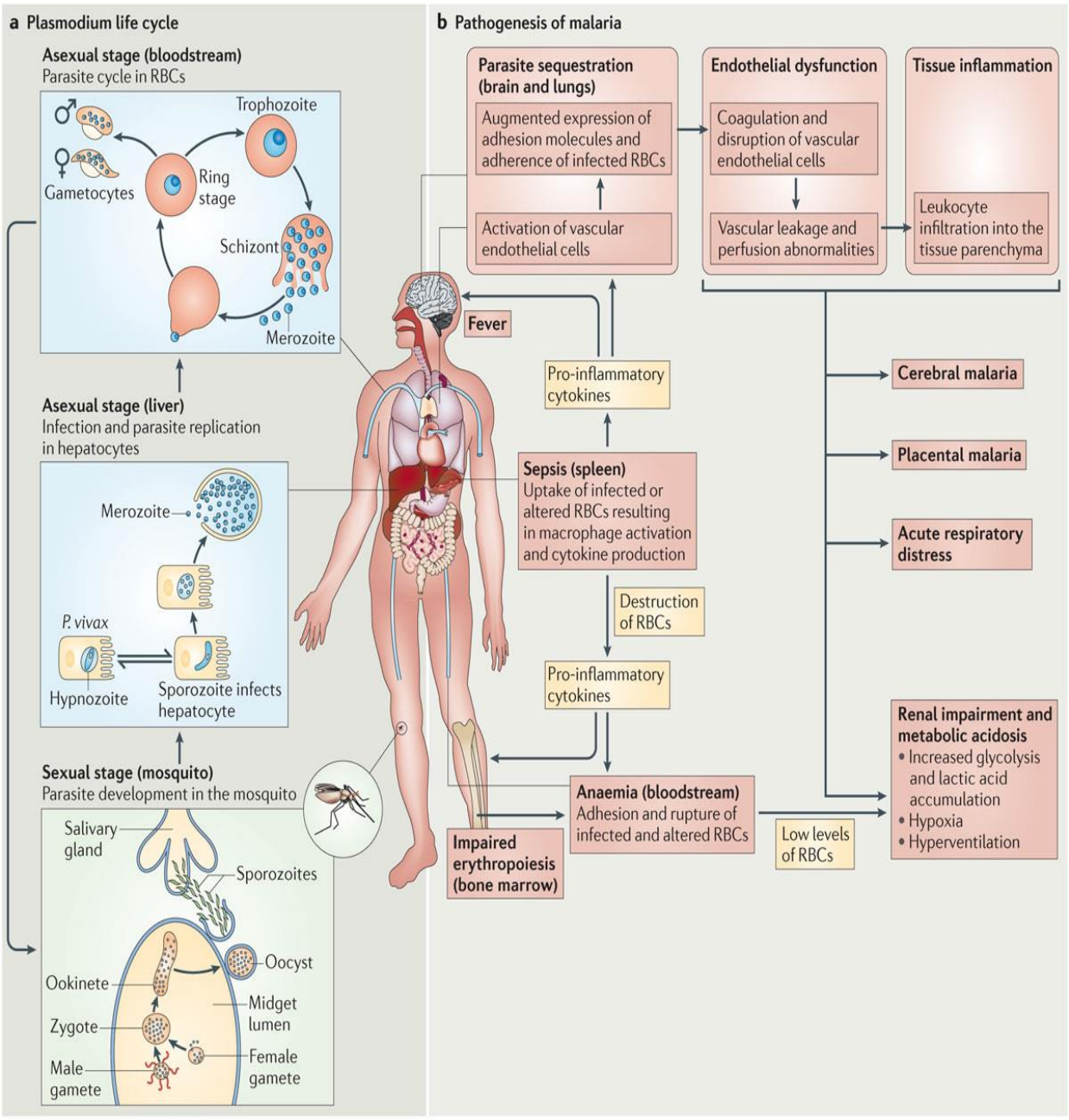


Figure 1: Pathogenesis of malaria. a) Different stages of asexual and sexual life cycle in mosquito and human host b) The pathogenesis and development of Malaria disease. Adapted from Gazzinelli et al [51].

1.5. Antimalarial agents and drug resistance

Antimalarials are the primary weapons in fighting against malarial disease since early times [107]. However, there are still limited number of existing antimalarials drugs on the markets. Here I discuss the most prominent antimalarial drugs and emergence and spread of drug resistance.

1.5.1. Quinine and Antifolate drugs

The bark of Cinchona trees known to contain several alkaloids was used as remedy for febrile illness among malaria endemic populations in Southeast Asia [108]. Quinine, one of the alkaloids was discovered in 1820 became the major drug until World War II when supplies of the drug from Southeast Asia was interrupted [108]. The discovery of quinine is considered the most successful medical discovery of the 17th century. Later efforts to prepare alternative synthetic antimalarial drugs led to the discovery of Chloroquine (CQ) in 1930s. Chloroquine became the most widely used synthetic antimalarial during the 1960s and 1970s [109]. Other common quinine-related compounds include amodiaquine (ADQ), primaquine, piperaquine (PIP) and mefloquine. Antimalarial antifolates have been central for prophylaxis and treatment of malaria [110] . Sulfadoxine-pyrimethamine was first introduced for treatment of malaria in Africa during the early 1980s for cases when chloroquine treatment failed, and has since become the first-line treatment in many countries. These drugs are usually administered as various combinations of dihydro-folate-reductase inhibitors (proguanil, chlorproguanil, pyrimethamine, and trimethoprim) and sulfa drugs (dapsone, sulfalene, sulfamethoxazole, sulfadoxine, and others) [110]. The combination includes sulfadoxine/pyrimethamine (Fansidar), sulfalene-pyrimethamine (metakelfin), and sulfamethoxazole-trimethoprim (co-trimoxazole). The antifolates act by inhibiting parasite enzymes necessary for folate biosynthesis and inhibit the synthesis of nucleic acids [110].

1.5.2. Artemisinin compounds

Artemisinin (Qinghaosu) was first isolated in 1971 by Tu Youyou from the plant Artemisia annua [111]. Tu Youyou received a Nobel Prize in Physiology and Medicine for her contribution in discovering Artemisinin [111] .These drugs demonstrated a very rapid clearance of the parasite from the circulation than any available antimalarials [110]. However, artemisinin has a very short elimination half-life and can't be used as malaria prophylaxis. Therefore, artemisinin is usually paired with other long acting drugs (Artemisinin-based combination therapy (ACT)) to allow parasite killing and protects against reinfection. Artemisinin-based combination therapy (ACT) is currently the mainstay of treatment for uncomplicated falciparum malaria in all malaria endemic regions (WHO). Commonly used ACT drugs includes: Artemether-Lumefantrine, Artesunate-Amodiaquine, Artesunate-Mefloquine, Dihydroartemisinin piperaquine and Sulphadoxine-Pyrimethamine [110].

1.5.3. Antimalarial drug resistance

Malaria treatment failure due to drug resistance is a recent phenomenon despite the long history of antimalarial use in endemic populations [109]. According WHO Antimalarial drug resistance is defined as "The ability of a parasite strain to survive and/or multiply despite the administration and absorption of a drug given in doses equal to or higher than those usually recommended but within tolerance of the subject" [110]. Factors that contribute to antimalarial drug resistance include widespread availability of low-quality medicines, non-compliance, incorrect use or suboptimum dose of drugs, nutritional status among others [110] .

1.5.4. Resistance to Chloroquine and sulfadoxine–pyrimethamine

QC resistant forms of *Plasmodium falciparum* malaria first appeared in Thailand in 1957 [109]. In 1970s, the emergence and widespread of QC resistance was reported in sub-Saharan Africa and South America which led to enormous human loss [112]. The spread of CQ resistance was a major factor for the failure of the first malaria eradication campaign in the middle of the 20th century [112]. QC is no longer effective against *P. falciparum*, but it still used for the treatment of *P. vivax* in regions where resistance has not developed [112]. QC acts by preventing detoxification of hematin within the parasite digestive vacuole [112]. Resistance to CQ is caused by the acquisition of multiple mutations in *P. falciparum* chloroquine resistance

transporter (*pfcrt*) protein encoded by the *pfcrt* gene on chromosome 7 [113]. *pfcrt* gene is believed to function as a transporter and is involved in the drug flux [114].

Sulfadoxine-pyrimethamine (SP) targets dihydrofolate reductase (PfDHFR) and dihydropteroate synthase (PfDHPS) which are enzymes in the folate biosynthetic pathway [115]. Accumulations of point mutations in these genes lead to resistance to SP; The more mutations accumulate in these genes, the greater the amount of resistance that is conferred to the parasite[116]. A similar pattern was observed with DHPS, with highly resistant lineages emerging in Southeast Asia and South America, and then spreading from Southeast Asia to Africa [116].

1.5.5. Resistance to Artemisinin

Artemisinin resistance is a major threat to global public health, with the most severe potential effects in sub-Saharan Africa where the malaria burden is the highest [117]. Resistance to Artemisinin (ART-R), characterized by delayed clearance of parasitaemia has emerged in South East Asia [118]. Resistance to Artemisinin has been associated with a mutation in a kelch protein located on *P. falciparum* chromosome 13 (K13 propeller) [119]. The K13-propeller mutation acts as a key causal determinant of artemisinin resistance has widespread distribution in the Greater Mekong subregion (GMS), which consists of Cambodia, Thailand, Vietnam, Myanmar and Laos [120].

In Africa, several mutations in K13 propeller mutations that have been observed at low levels [121]. However, these mutations are different from the Mutation in SEA and not associated with delayed clearance of parasites [121]. However, a recent study in Rwanda reported the existence of *Pfkelch13 561H,* one of the K13 mutations in SEA and its clonal expansion. The authors further validated as a mediator of ART-R in vitro [122]. The findings of this study have substantial implications for public health in confirming the de novo emergence and clonal expansion of an ART-R *Pfkelch13* R561H lineage in Rwanda which can potentially compromise the success of malaria control in Africa [122]. Therefore, it is imperative to undertake continuous genetic surveillance and monitoring of antimalarial drug resistance in endemic regions. Furthermore, alternative antimalarial drugs are urgently needed to realize the global malaria eradication.

1.6. Malaria vaccine strategies: progress and limitations

The field of malaria vaccine has faced many challenges including the genetic and biological complexities of the *Plasmodium* parasite, presence of multiple, antigenically distinct life cycle stages in the mammalian host, the involvement of mosquito vectors and the prevalence of malaria in the in resource constraint populations among others [123] . The WHO set a malaria vaccine development roadmap to be met by 2030: these include (1) developing vaccines with protective efficacy of at least 75% against clinical malaria and, (2) vaccines that reduce transmission of the parasite to reduce the incidence of human malaria infection [123] . Several strategies have been implemented to develop vaccine that can be used in mass vaccination programs to facilitate the global malaria eradication efforts [124]. These including pre-erythrocytic, blood-stage and mosquito-sexual stage of development [125]. Here I discuss the recent progresses in malaria vaccine developments and their limitations.

1.6.1. ***Pre-erythrocytic vaccines***

The Pre-erythrocyte (PEV) vaccine approach aim to target the malaria parasite during its sporozoite and liver stage of development and hence avert the progression to the blood-stage [125]. PEVs are designed to induce (1) antibodies against surface antigens that clear sporozoites from skin or bloodstream or block their invasion of hepatocytes, or (2) T cell responses that attack infected hepatocytes [125]. The two most successful Pre-erythrocyte vaccine thus far include whole organism sporozoite vaccine and a recombinant protein in adjuvant vaccine called RTS, S vaccine.

1.6.1.1. RTS, S

RTS, S, is the first human parasite vaccine passed the highest level of the WHO regulatory scrutiny in 2015 [125]. It is composed of the major surface antigen of sporozoites (CSP): fragment comprising central repeat (hence "R") and C- terminal regions (containing T cell epitopes, hence "T") fused to hepatitis B surface antigen ("S"). RTS is expressed in yeast that also carry hepatitis B "S" expression cassettes, and thus synthesize S and hence given the name RTS, S [126]. Recommended by the WHO, RTS, S/AS01E pilot implementation programs were launched in several places 2019 to assess safety and benefits during delivery through

standard public health mechanisms [126]. However, the protection offered by RTS, S/AS01 vaccine is modest and wanes over time [126]. R21 is an improved version of the RTS, S/AS01 vaccine designed by incorporating a higher proportion of *Pf* CSP C-terminus bound to HBsAg N-terminus without the three-fold molar excess of HBsAg found in RTS, S/AS01. This enhanced B cell activation led to stronger anti-CSP humoral immune responses was immunogenic even at very low doses in mice and has been under evaluation in Phase 1/2a clinical trials [127]

1.6.1.2. ***Whole Sporozoite Vaccines***

This approach was introduced in 1970s and has been formulated in different ways including radiation-attenuated Sporozoite (Spz), genetically-attenuated parasite and Spz administered together with antimalarials [125]. Implementing these vaccines has been practically difficult because of the challenges including the requirement of liquid nitrogen cold chain, intravenous inoculation, difficulty to scale-up of the vaccine manufacture [125]. In effort to address these challenges, Sanaria company introduced a technology that is capable of harvesting PfSPZ from the salivary glands of aseptic mosquitoes infected by cultured laboratory parasites, followed by purification, vialing, and cryopreservation in liquid nitrogen vapor phase in 2010 [125].

A recent study in Mali showed that the PfSPZ vaccine is well-tolerated and have 29% efficacy against heterologous strains [128]. Many studies have been aimed at improving attenuated Spz vaccines, focusing on efforts at producing a large repertoire of immunogens, evaluating the impact of a particular regime, dosage and inoculation route, thereby enabling an effective cellular and humoral immune response to be achieved[129]. Future improvements are needed to broaden efficacy against heterogeneous parasites.

1.6.2. Blood-stage vaccines

Blood-stage malaria vaccines (BSV) are attractive in that they target the disease-causing stage of the parasite development [125]. It has long been shown that passive transfer of IgG purified from semi-immune African adults were able to clear parasitaemia from African children and later in Thai adults [124]. Several studies have been conducted to produce effective BSV vaccines the majority of which target the merozoite surface protein (MSPs), apical membrane antigen1 (AMA1) and *P. falciparum* reticulocyte-binding protein homolog 5 (PfRH5) [130]. Generally, these studies have shown that BSV vaccines can illicit some amount of that would

impair parasite invasion and infection in laboratory environment or experimental animals. However, the results showed scant evidence of protection against controlled human infection or against naturally occurring infection [125]. Among BSV candidates, only GMZ2 (consisting of conserved domains of GLURP and MSP3) showed statistically significant (14%) efficacy in Africa [131].

The challenges of developing effective anti-merozoite vaccines include (1) the very short time when merozoites are accessible to antibodies, (2) antigenic polymorphism, (3) redundant invasion pathways, and (4) the large number of parasites that need to be targeted [125].

1.6.3. ***Transmission-blocking vaccines***

Transmission-blocking vaccine (TBVs) target the sexual stage of malaria parasite in mosquitoes host and hence block transmission of malaria[132]. Effective TBVs would ideally be used to protect the immediate neighborhood of the vaccinated individuals if high population coverage is achieved [125].

The leading candidates are grouped as gamete surface proteins first expressed by gametocytes in human blood such as Pfs230 and Pfs48/45 of *P. falciparum (BVS1 et al)*, and zygote surface proteins expressed only post-fertilization in the mosquito host such as Pfs25 and Pfs28 [133]. However, because these vaccines do not prevent infection in vaccinated individuals they are an unattractive financial venture for vaccine companies and thus far, demonstrated only modest efficacy [134]. The challenges of TBV development include achieving sufficient adaptive responses that maintain high levels of antibodies over time, widespread coverage to accomplish herd immunity and adequate safety profile to immunized individuals[134].

"The corpuscles of anaemic heterozygotes are smaller than normal, and more resistant to hypertonic solutions. It is at least conceivable that they are more resistant to attacks by the protozoa which cause malaria, a disease prevalent in Italy, Sicily and Greece, where the gene is frequent" (Haldane, 1949).

1.7. Human genetic polymorphisms and severe malaria: From few genes to whole genome

A comprehensive understanding of the genetic basis of severe malaria resistance can potentially inform the development of treatments and vaccines [135] . A classic example is the success achieved in the development of candidate vaccine for *P. vivax,* the predominant causative agent of malaria in South America and Asia [136]. The fact that this parasite is remarkably rare in sub-Saharan Africa had long been a puzzle for scientists [137]. After a series of studies, it was revealed that most Africans lack the Duffy blood group, an essential receptor for erythrocyte invasion by *P. vivax* [137]. Later, the discovery of causative SNP in the chemokine receptor (DARC) gene and subsequent molecular analysis of this genomic region led to the development of candidate vaccine [138]. Continuing efforts have been made to make useful genetic discoveries for *P. falciparum* for the last several decades [135].

P. falciparum malaria is one of the best-known strong evolutionary selective forces acted on the human genome [139]. Malaria parasites have been infecting humans for at least 5000-10,000 years following the emergence and expansions of agriculture [139]. The discovery of major erythrocyte disorders (hemoglobinopathies) in a higher frequency in some populations puzzled the geneticists of 20th Century [140]. The majority of the scientists of the time associated this phenomenon with human mutations as a consequence of Second World War [141]. However, J.B.S.Haldane, one of the most distinguished founders of population genetics, was unsatisfied with this argument and suggested that "The corpuscles of anaemic heterozygotes are smaller than normal, and more resistant to hypertonic solutions. It is at least conceivable that they are more resistant to attacks by the protozoa which cause malaria, a disease prevalent in Italy, Sicily and Greece, where the gene is frequent" [142]. This insight soon became the so called "Malaria hypothesis" that propose inherited red cell disorders including thalassaemia and sickle cells are the results of natural selections by malaria [141]. Even though malaria hypothesis was first developed with

respect to the high frequency of thalassaemia genes, it turned out to be true for sickle cell trait and other hemoglobinopathies [143].

Hemoglobinopathies are caused by structural and functional defects of human haemoglobin and are distributed in malaria endemic countries or in areas where malaria was prevalent [141]. Human haemoglobin is a tetrameric globular protein consisting of two α and two β polypeptide chains each having an oxygen-binding heme [144]. The α-chain (*HBA*) has 141 amino acids and the β chain (*HBB*) has 146 amino acids [144]. The genes that code for α chain are located in the telomeric region of the short arm of chromosome 16 (16 p 13.3), whereas the genes encoding for the β chain are located on the short arm of chromosome 11(11p.15.5) [144]. Among several structural Hb variants discovered so far, only three *HbC* (β 6Glu→Lys), *HbS* (β 6Glu→Val), and *HbE* (β 26Glu→Lys) reached polymorphic frequencies; suggesting that mutants at other sites of *HBB* locus and all sites at *HBA* loci, are either neutral or detrimental in malaria endemic environments [145].

The majority of hemoglobinopathies are related to "loss of function" resulting in reduced expression or altered gene products [145]. As a result, inherited red blood cell disorders are among the most common genetic diseases in humans (more than 300,000 children are born each year with a severe inherited haemoglobin disorders) [145]. Furthermore, it has also been suggested that immunological, inflammatory and other chronic diseases prevalent today may be attributed to the pleiotropic effects of mutations that confer malaria resistance [146]. Therefore, population genetic studies of malaria resistance have huge implications in human health and evolutions. It is now generally agreed that *P. falciparum* malaria caused a number of host-protective polymorphisms at high frequencies because of its ability to kill before reproductive age [147]. Next, I discuss the well-known variants. I will discuss the novel variants identified by GWASs in Chapter 2.

1.7.1. Haemoglobin S (HbS)

Sickle cell trait has been recognized as the most prominent human genetic factor that confer protection against the lethal form of malaria [141]. It is caused by a SNP (*rs334*) in the coding region of *HBB* on chromosome 11p15.4 which results in substitution of Glutamic acid with Valine at amino-acid residue 6 of the β -globin chain [143]. Unlike homozygous (*SS)* individuals who suffer from a severe haemolytic syndrome, heterozygous (*AS)* are healthy and are protected from the severe form of malaria [140].

Haplotype studies suggested that the mutation of *HbS* independently arose in different populations at least one in Africa and one outside Africa [148]. Within Africa, the presence of the same mutation on different haplotype in different populations may suggest that the independent mutation likely arose more than once in this continent [149,150]. However, it is also possible that haplotype diversity can be caused by the differences in gene background conversion and recombination in different populations [149]. *AS* has shown to have a protection of approximately 90% against SM [151]. *HbS* is widespread all over malaria endemic areas and common in sub-Saharan Africa with a prevalence reaching 15-20% [140].

1.7.2. Hemoglobin C (HbC) and Hemoglobin E (HbE)

HbC is caused by change in b-globin molecule at the same amino-acid position as for *HbS*, but the change from Glutamic acid to Lysine rather than to Valine [152]. While homozygous form (CC) causes mild symptom of chronic haemolytic anaemia, the heterozygous (AC) is asymptomatic [153]. Both heterozygous (AC) and homozygous (CC) form confer protection against SM [154]. This may suggest that *HbC* could eventually reaches fixation in malaria environment with better fitness compared to *HbS* [154]. However, the current distribution of *HbC* is mainly limited to west Africa compared to the wider distribution of *HbS* [155]. *HbE* is caused by substitution of Lysine with Glutamic acid in codon 26 [156]. This structural variant is common in South East Asia; reaching up to 70% in northern Thailand and Cambodia [140]. Linkage disequilibrium analyses suggest that *HbE* variant was emerged recently and come under intense selection in South East Asia [157]. The heterozygous state (*AE)* confers protection against SM [157].

1.7.3. Thalassemia

α-thalassemia has two major varieties including α^0-thalassaemia, in which both of the linked *HBA* genes are deleted, and α^+-thalassemia in which one of the pairs of linked genes (*HBA1* and *HBA2*) on chromosome 16 is deleted or inactivated by a point mutation [140]. The homozygous state is represented by --/-- and–α-α for α^0-thalassaemia and α+ thalassaemia, respectively [140]. While α^+ thalassemia cause a mild hypochromic anaemia in their homozygous state, homozygous form of the α^0 (--/--) is a lethal disease resulting in stillborn and foetal death [140]. The distribution of α-thalassemia variants overlaps with malaria

distribution in past or present including sub-Saharan Africa and the Mediterranean region, the Middle East, the Indian subcontinent and south East Asia [140]. Homozygous state and heterozygous state of α^{+}-thalassemia confers an estimated protective level of 63% and 83%, respectively [151].

β-thalassemia is caused by the reduction or loss of functional protein in *HBB* gene on chromosome 11 [158]. More than 200 mutations have been reported; the large majorities are point mutations in functionally important regions of β-globin genes [158]. The frequency of β-thalassemia variants ranges from 5-20% in global population [140]. The severity of the disease varies depending on the nature of β-thalassemia allele, but at its most extreme form can resemble sickle-cell anaemia [159]. Individuals with β-thalassemia major (homozygous for the loss of production of β-chain) have profound anaemia and if not treated with blood transfusion, die in their first year of life [160]. But heterozygous individuals (β-thalassemia minor) have only mild anaemia [160]. β-thalassaemia trait has been reported to play protective role against SM [161].

1.7.4. Proposed protective mechanisms hemoglobinopathies

The molecular mechanisms that confer protection against severe malaria remains incompletely understood [151]. Nevertheless, it has been showed that the protection is attributable to genetic, molecular and immunological factors in which the parasite growth is impaired or its removal is enhanced [162]. Studies showed that ring-stage parasite don't grow under low oxygen tension in *HbS* red cells [163,164]; suggesting that hemoglobinopathies interfere with intra-erythrocyte growth of the parasite. Another interesting invitro study implicated miRNA for the parasite growth inhibition in RBCs [165]. The authors observed translocation and fusion of several host miRNAs with parasite's mRNA transcripts in in RBCs collected from *AS* and *SS* individuals. The fusion of the miRNAs was shown to inhibit the translation of enzymes that are important for the parasite development [165].

In addition to interference with the parasite invasion and intraerythrocytic development, hemoglobinopathies have shown to neutralize the pathogenic pathways of the SM [151]. For instance, *HbS* and α^{+}-thalassemia impair cytoadhesion and rosetting which are the major mechanisms of pathogenesis [61]. *HbC* and *HbS* have been shown to reduce parasite induced remodelling of iRBCs and thereby reduce adhesion properties of the iRBCs [166]. It has also been shown that the rate of sickling is higher in parasitized RBCs compared to the non-

parasitized RBCs in HbAS individuals [167]. Sickling RBCs are rapidly removed from the circulation by immune system and hence minimize the chance of developing complications that lead to severe malarial disease [167]. The detailed proposed mechanisms of protections were described in [151].

1.7.5. Enzymopathies: Glocose-6-phosphate Dehydrogenase (G6PD) and Pyruvate kinase

G6PD is an important enzyme that catalyses the first reaction in the pentose phosphate pathways [168]. It is a key enzyme to produce reduced glutathione and control oxidative stress in cell [168]. *G6PD* deficiency is one of the most common inherited disorders affecting around 400 million people worldwide with a global prevalence of 4.9% [169]. The *G6PD* gene is located on the telomeric region of the long arm of X-chromosome (Xq28) close to the genes for haemophilia and colour blindness and is 18kb long consisting of 13 exons [168]. The region that encodes *G6PD* gene is one of the most polymorphic loci in human genome in which about 140 different variants have been identified [168,170].

The global prevalence of *G6PD* deficiency is geographically correlated with areas inhabited by populations historically exposed to malaria including Africa, Mediterranean, South East Asia and Latin America [170]. In Africa, the most frequent variant is a mutant allele A- caused by nucleotide changes either at A376 (Asn to Asp) or changes at G202A (Val to Met) [170]. It has been estimated that *G6PD* A- allele arose within the past 3,840-11,760 years and spread with expansion of malaria [170]. Some studies reported that both hemizygous males and homozygous females are protective against SM [170]. However, a relatively more comprehensive study showed that *G6PD* deficiency protects hemizygote males but not heterozygous females from SM [171].

Deficiency of pyruvate kinase, a rate limiting glycolytic enzyme, has also been implicated in conferring protection against malaria [149]. The enzyme is encoded by *PKLR* located on chromosome 1 (1*q22)* and abundantly expressed in RBCs and Liver [172]. Mutations in the isoform which is exclusively present in RBCs causes an autosomal recessive inherited pyruvate kinase deficiency which results in a nonspherocytic haemolytic anaemia [172]. Genetic studies in mouse models [173] and invitro analysis in humans [174] suggested that the pyruvate kinase deficiency confer protection against malaria.

1.7.6. Erythrocyte membrane polymorphisms: ABO blood group, Ovalocytosis and Glycophorins

The histo-blood group *ABO* consists of three carbohydrate antigens A, B and H [175]. A, B and AB express glycosyl-transferase that adds either N-acetylgalactosamine to form the A antigen, galactose to form the B antigen, or leaves the H (O) antigen unmodified [176]. The *ABO* gene is located on chromosome 9q34, and the A and B alleles are co-dominant against the recessive O allele [175,177]. Similar to hemoglobinopathies, loss of functional allele in *ABO* gene (allele O) provides survival advantage while group A confers disadvantage and group B has intermediate effect against SM [178]. The prevalence of group O higher in malaria endemic regions compared to group A [178].

Ovalocytosis is an autosomal dominant hereditary condition that affects erythrocyte membrane rigidity and shape [179]. The term 'ovalo' is referring to the oval shape of erythrocytes of affected individuals [179]. Ovalocytosis is caused by a 27 bp deletion in *SLC4A1* gene on 17q21.31 which encode 'band 3' protein [180]. 'band 3' is a major transporter of anions across the erythrocyte membrane and also involves in maintaining integrity of the erythrocyte membrane [180]. Ovalocytosis is highly prevalent in parts of South-East Asia and its distribution coincides with malaria endemicity. Ovalocytic erythrocyte partially resists invasions by both *P. vivax* and *P. falciparum* and confers protections against malaria [181]. It has also been reported that Ovalosytosis confers protection against CM [182].

Another erythrocyte membrane component implicated in malaria resistance are Glycophorins gene cluster including *GYPA*, *GYPB* and *GYPC* [183]. Glycophorins constitute the sialoglycoproteins of erythrocyte membrane which determine different blood group antigens and are recognized as receptors of *P. falciparum* merozoites [184,185]. *GYPA* and *GYPB* acts as a receptor for *P. falciparum* erythrocyte-binding antigen 175 (*EBA 175*) and erythrocyte-binding ligand 1 (*EBL-1*), respectively and determine MNS blood groups [184,185]. *GYPC* determines Gerbich blood group and serves as receptor for the *P. falciparum* erythrocyte-binding antigen 140 (*EBA140*) [186]. Glycophorin variants have been shown confer malaria protection by interfering with the invasion pathways of the parasite [97].

1.7.7. Haptoglobin (Hp)

Haptoglobin (Hp) is an acute phase glycoprotein present in human plasma encoded by a gene on 16q22 [187]. There are three genetic variants of Hp in humans including *Hp1, Hp2* and *Hp2-1* [187]. *Hb* is functions as the main scavenger of toxic free hemoglobin (Hb) resulting from intravascular haemolysis. *Hp* binds to extravascular *Hb* and forms Hp-Hb complex which is removed by a CD163, a receptor, abundantly expressed on macrophages [188]. In SM, *Hp* forms stable complexes with extracellular *Hb* that is released from parasite driven-lysis of RBCs and thereby curtail the haemoglobin-induced oxidative tissue damage [189]. However, the protective role of different Hp genotypes on SM remains controversial as reviewed in [149].

1.7.8. Polymorphisms in adhesion molecules: CD36, ICAM1and CR1

P. falciparum differs from other species in that the infected red blood cells (RBCs) adhere to vascular endothelium and disappear from the blood circulation for the entire life cycle to avoid splenic clearance and host immunity [58]. *CD36* is one of the major host receptors to *CIDRa2-6 domain* of the parasite's *PfEMP1* [190]. *CD36* is a membrane glycoprotein encoded by a gene on chromosome 7q21.11 and present on platelets, mononuclear phagocytes, adipocytes, hepatocytes, myocytes [191]. It is a cell surface receptor which plays a role angiogenesis and phagocytosis [191]. It has been shown a missense mutation of the gene encoding of *CD36* confers protection against SM by reducing parasite sequestration [192].

ICAM1 is an inducible glycoprotein belonging to immunoglobulin superfamily and is encoded by gene on chromosome 19p13.2 [193]. It another receptor for *PfEMP1* which is widely expressed in brain tissue (CM) [72]. A polymorphism caused by A-to-T transversion at nucleotide 179 of N-terminal domain of *ICAM1* (also called *ICAM1 Kilifi*) has shown to be linked to have a high frequency in malaria endemic populations [194]. *ICAM1 Kilifi* has been associated with increased susceptibility to SM in Kenyan population. The same mutation has been reported to be associated with resistance against SM in Gabon population [195]. However, another study conducted Malawi, Kenya and Gambian populations reported that there is no association between SM and variants of *ICAM1* gene [196].

CR1 is an immune-regulatory protein encoded by a gene located on chromosome 1q32.2 and is involved in complement activation and clearance of immune complexes [197]. *CR1* is abundantly expressed in RBCs and leukocytes and has been shown to be associated with rosetting, one of the major factors that lead to sequestration and associated pathologies during SM [58]. Deficiency of *CR1* has been shown to be common in malaria endemic populations in Melanesian populations [198]. The same study demonstrated that *CR1* deficient individuals are protected against SM [198]. Further studies have reported that polymorphisms in CR1 genes confer protection in Kenyan [199] and west African populations [200].

*1.7.9. Polymorphisms in immune response genes: HLA, IFN*γ

Human Leukocyte Antigen (*HLA)* is encoded by the Major Histocompatibility Complex (MHC) which is the most polymorphic genes known in human genome. The diversity of MHC is believed to be driven by selection pressure from infectious pathogens and known to be associated with the risk of several infectious diseases [201]. *HLA* variants such as *HLA* class I antigen (HLA-Bw53) and *HLA* class II (DRB1*1302-DQB1*0501) were reported to common in African populations and confer some degree of protection against severe malaria [202]. *HLA* class I antigen is expressed by liver cells and hence, T cells (CTL) responses might more efficiently act against the liver stage of malaria parasite in individuals with HLA-Bw53 [202]. Whereas, individuals with DRB1*1302-DQB1*0501 variant might possess efficient antigen presentation mechanism that can lead to rapid clearance of blood stage parasite [202].

Massive release of pro-inflammatory cytokines is one of the factors that lead to malaria pathogenesis. *IFN*γ is encoded by a gene on 12q15 and plays a regulatory roles in development and functions of immune system[203]. It is one of the major cytokines, abundantly during malaria infection caused be deficiency caused by a polymorphism at promoter region of this gene (*IFN*γ 183G/ T) has been associated with protection against CM in Malian populations[204]. Moreover, polymorphisms in transcription factor interferon regulatory factor 1 (IRF-1) which regulates *IFN*γ production is shown to confer protective role against SM [205]. Similarly, polymorphisms of *IFN*γR1 gene have been reported to confer resistance against CM in Gambian populations [78]. Several other immune variants were reported to be associated with severe malaria resistance trait. However, only few were replicated across in different populations.

1.8. Approaches to study human genetic resistance and susceptibility to severe malaria

1.8.1. Candidate gene association studies

Candidate gene association study approaches is one of the best known genetic epidemiological methods used to establish relationship between genes and diseases [206]. It is a knowledge based, hypothesis-driven approach which enables to test associations between a specific genomic regions and the phenotype of interest [206]. Candidate gene association study approaches include cohort studies and case control studies. In cohort studies, individuals who have and do not have specific alleles or genotypes grouped in separate cohorts and followed prospectively to determine whether exposure affects the risk of developing diseases. The comparison is made by calculating relative risk between the groups. In case-control studies, the study individuals are grouped based on disease status and examined for exposure retrospectively. The case-control study approaches compare the frequencies of genetic markers in cases and controls. The comparison is usually presented by calculating the odds ratio between the cases and controls [206].

Candidate gene association study approaches have been successfully applied to malaria susceptibility/resistance studies for more than two decades and identified a number of genes [141,147,149]. However, several conflicting findings were reported from different studies [149]. Some variants that were reported to be associated with protection against SM in one population were found neutral in another populations [141]. In some instances, variants that were associated with the protection against SM in initial population were found to be associated with increased risk in the second population when replication was attempted [149]. Consequently, only few of the associations were replicated in different populations. For instance, in a multi-center case-control association study conducted in malaria endemic countries, forty percent of the previously reported loci (22 out of 55) were failed to replicate [207].

The major limitations of this approach is that the candidate genes are selected based on prior knowledge of biology and epidemiology of malaria; specifically, the observation of balancing selections of inherited red cell disorders[147]. As a result, the majority of the identified risk loci are erythrocyte variants that cause red blood cell diseases as reviewed elsewhere [141]. Erythrocyte variants can easily be detected compared to the phenotypic consequences of malaria in other cells. For instance, alteration in liver cell receptor is difficult to observe

compared to a sickling red cell [147]. Thus, other potential genes that influence malaria resistance/susceptibility might be missed. Another limitation of the conventional candidate gene approach is that, it can't accommodate polygenic genetic architecture in which multiple variants with small effects add up to control the phenotypic variance[207].

1.8.2. Family based linkage studies

Linkage study approaches aim to identify genetic associated loci using families or set of families that display sufficient phenotypic variations for the disease/trait under investigation [208]. The co-segregations of the genomic region that contain causal variant/variants within the family are assessed by set of selected polymorphic markers. Linkage study approaches can be used to generate hypothesis because, it enables identifying the genomic regions of association without the need of previous knowledge about the link between the certain genomic regions and the disease of interest [208]. The result of linkage analysis is usually reported as a logarithm of the odds (LOD) score.

Linkage studies have been applied in malaria as a classical approach to identify genes that are co-segregating in families [209]. Several genomic loci including 1p36, 2p25,4q13,5q31-q33, 6p25,9q34,10p15.3, 10p14, 12q21-q22,13q13,20p12 and 20q11 were reported to be associated with malaria infection and severity in different African populations [209–214]). However, the majority of the findings were discordant and failed to be replicated in different populations.

This is mainly due to the fact that the sample sizes in linkage studies are generally small and the variants assayed are limited to several hundred to thousands of markers which lacks the resolution power to detect genes with small effect sizes [215]. Moreover, the markers are usually heterogeneous in populations with different genetic backgrounds; implying that association signals can be dissolved during the analysis and leads to false negative results.

1.8.3. Genome-wide association studies (GWAS) and Post-GWAS

Genome wide association studies (GWAS) approach has been proven to be a powerful tool for investigating the genetic basis of common diseases [135]. This approach overcomes the majority of the fore-mentioned challenges in candidate gene and linkage studies and provided unprecedented successes in genetic studies of a wide range of complex-traits and common diseases [135,215]. The fundamental rationale of GWAS is a 'common disease common variant' hypothesis which predicts that common diseases are influenced by common genetic

variants in populations [216]. GWAS makes use of commercial 'SNP chips' that are designed to capture the majority of the common variations in human genome. The representative SNPs (tag SNPs) in genotyping array are selected based on the linkage disequilibrium (LD), the correlation between neighbouring SNPs in populations because of past evolutionary history [216]. Following the availability of comprehensive reference data and SNP catalogue, denser commercial 'SNP chips' become available for GWAS; some of which currently contains up to 2,000,000 SNPs [217]. This further facilitated the application of GWAS in a wide range of genetic studies of complex-traits and diseases [217].

The application of GWAS approach to malaria research was motivated by two main reasons: (1) Malaria has been a strong selective force in the recent human evolution; 2) Only a small proportion of heritability of malaria resistance trait is explained by the well-known variants [135,218]. The Malaria Genomic Epidemiology Network (MAlariaGEN) [20], a global partnership of malaria researchers, has been established to conduct multi-centre-scale genetic-association studies of malaria resistance and susceptibility. The consortium addressed the major practical and ethical obstacles including recruiting large number of patients, implementation of data management and sharing systems, proper classification of severe malaria phenotype, standardization of sampling and sample processing systems across the study sites [20].

This led to the successful implementation of several GWASs in malaria endemic regions [150,219–223]. Importantly, a central repository of genotypic and phenotypic data was established which researchers can access up on acceptance of their proposals and signing of a legally binding data access agreement (www.malariagen.net/resourse/2) to facilitate further scientific discoveries [224]. In addition to GWASs, recent advances in post-GWAS including capturing polygenicity, risk prediction, detecting new risk loci, imputing un-typed associated variants and fine-mapping causal variants from GWAS summary statistics and LD information based on population specific reference panel are also playing an increasingly critical role in complex disease studies [217]. However, the severe malaria resistance GWASs have been suffering from several challenges. I will discuss the progress and pitfalls of the application of GWAS approach in MalariaGen dataset in chapter 2

1.9. Motivations of the project

The global malaria eradication efforts have been implemented in malaria endemic areas using conventional strategies including distribution of long-lasting insecticide treated nets (LLINS), indoor residual insecticide spraying, intermittent treatment for pregnant women in high transmission settings [11]. This led to the significant decline of malaria burden in many parts of the endemic regions [12]. Despite the successes gained, the progress towards global malaria elimination is currently challenged by emergence of drug resistant parasites, insecticide resistant mosquitoes and lack of effective vaccine [12]. In effort to devise host-directed intervention strategies such as therapeutics and vaccine, several GWASs have been implemented in the last decade and shed new light to the genetic bases of malaria resistance.

Even though the GWASs enabled better understanding of the genetic basis of the complex disease/trait, the method suffers from the following shortcomings:1) the weak performance in genetically diverse populations; 2) the lack of translation of associated loci into suitable biological hypotheses 3) the well-known problem of missing heritability; 4) the lack of understanding of how multiple modestly associated loci within genes interact to influence a phenotype; 5) inefficiency in distinguishing between inflation from bias (cryptic relatedness and population stratification) and true signal from polygenicity; 6) the imperfection of asymptotic distribution of current mixed model association or logistic regression in the specific case of low-frequency variants. These shortcomings were clearly reflected in previous severe malaria resistance GWASs [150,219,220,223].

For instance, in one of the studies, HbS locus, a well-known variant conferring resistance to severe malaria demonstrated a weak signal (p values ~ 10^{-7}); partly because of the weak LD between causal variants and the SNPs that were genotyped, suggesting that several other signals with modest effect sizes were missed. On the other hand, several genomic regions containing variants with evidences of associations were identified. For instance, in malaria GWAS studies conducted in eleven populations, 34 regions of the genome that contain variants with evidence of associations with severe malaria were identified [219]. Earlier GWAS in Ghanaian population also identified 40 genomic regions containing 102 SNPs with evidences of association with protection against severe malaria in the discovery phase of the

study [225]. However, the majorities of the variants fail to pass the stringent significance GWAS threshold ($p=5x10^{-8}$) required to report genuine associations.

This raises several questions including 1) What is the genetic architecture of malaria protection? 2) What is the contribution of polygenic effects and their distributions across cell-types, chromosomes, functional groups and molecular pathways? 3) What is the extent and pattern of epistasis and pleiotropy at genome wide scale which are known to influence malaria resistance trait among others. This information has a huge potential to allow us understand the underpinning biology and hence paves ways to design alternative therapeutics and vaccine candidates. Another obstacle is that the genetic determinants of malaria resistance trait and their effect sizes have been shown to vary between different African populations in previous GWAS studies. This could be because of several reasons including differing environmental factors, differing patterns of linkage disequilibrium (LD) and differences in allelic architecture due to the population diversity in Africa. Furthermore, the biological and functional information of the identified variants have not been well elucidated at gene and pathway levels.

In attempt to address the fore-mentioned challenges, advanced statistical approaches that can improve power of association studies have been recently developed and implemented in complex disease studies [226–229]. In addition to this, recent advances in post-GWAS methods enabled applications of gene-based GWASs, functional studies, protein-protein interactions, and polygenic analysis on publicly available GWAS-summary statistics without the need of individual genotype datasets [230–234]. However, the majority of these approaches have not been implemented in malaria resistance GWAS datasets. Given the availability of malaria resistance GWAS datasets and improved public reference panels, it is now an intriguing time to investigate the genetic basis and underpinning biological functions of malaria resistance trait. We therefore, aimed to implement contemporary statistical genetic analytic approaches to malaria GWAS datasets to dissect the genetic basis of malaria resistance trait.

1.10. Objectives of the thesis

1. To explore the challenges of GWAS approaches in severe malaria datasets and figure out how various advanced statistical genetic methods can be implemented to address these challenges.
2. To estimate the SNP heritability of malaria resistance in African population and partition in to chromosomes, allele frequencies and annotations
3. To prepare African specific reference panel from public datasets to enable LD-based enrichment analysis from summary statistics
4. To identify candidate malaria resistance genes and pathways
5. To compare population genetic structure of candidate malaria resistance gene in malaria endemic populations and global populations
6. To identify rare variants that influence malaria resistance trait

1.11. Outlines of the thesis

In CHAPTER 1 of this thesis, we presented a general introduction as described above. In CHAPTER 2, we discussed progresses and challenges of the conventional GWAS approaches in African populations in general and in malaria resistance GWAS datasets in particular. We presented alternative strategies that can potentially be applied to the GWAS datasets and provide better insights to the genetic basis of severe malaria resistance. In CHAPTER 3, we presented the results from comprehensive heritability analyses of severe malaria resistance in three African populations including Kenya, Gambia and Malawi. As part of enrichment analysis, we prepared African specific reference panel obtained from African populations datasets in 1000 Genomes Project and African Genome Variation Project reference datasets and computed LD for the panels. In CHAPTER 4, we presented the results obtained from various functional analyses performed on malaria resistance GWAS dataset of 17,000 individuals meta-analysed across eleven populations in Africa, Asia and Oceania. We further presented the findings obtained from population structure analyses and rare variant analysis performed on raw genotype GWAS datasets (N = ~11,000) of Kenya, Gambia and Malawi populations. In CHAPTER 5, we summarized all the findings and discussed the future works that need to be done to provide better insights in to genetic basis of malaria resistance.

1.12. Overall contributions of thesis project in malaria field

In this work, by noticing the fact that performance of conventional GWASs is relatively weak in African populations in general and in malaria resistance GWAS datasets in particular, we figured out how advanced statistical genetic analytic approaches can be applied improve the power of these studies. We showed that the statistical genetic methods that have been developed for homogeneous populations such as Europeans can be applied to the genetically diverse African dataset when population specific reference panels are used and stringent quality control strategies are implemented. The reference panel we prepared as part of our analyses can potentially be used for other LD-based studies in African populations.

We showed for that malaria resistance is polygenic trait with SNP-heritability (h^2_g) of ~20% and that the causal variants are overrepresented around protein coding regions of the genome.

By applying several gene-based, pathway-level and network-level functional analyses to severe malaria resistance GWAS summary statistics (N=17,000) meta-analysed across eleven populations in malaria endemic regions in Africa, Asia and Oceania, we systematically identified 57 genes located in the known malaria genomic loci and additional 125 genes across the genome.

We showed that identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signalling elements and neuronal systems. Overall, our works showed that severe malaria resistance trait is attributed to multiple genes that are enriched in pathways linked to severe malaria pathogenesis. These findings laid the foundations of future experimental studies which can potentially lead to translational medicine including the development of vaccine and new therapeutics.

2. CHAPTER TWO: GENOME-WIDE ASSOCIATION STUDIES OF SEVERE *P. FALCIPARUM* MALARIA SUSCEPTIBILITY: PROGRESS, PITFALLS AND PROSPECTS

Summary

In this chapter, we assessed the progress of severe malaria resistance GWASs, presented the biology of the novel variants identified by the previous GWASs studies and proposed the mechanisms of protections of these variants against severe malaria. We extended our discussion to the challenges that compromise the power of the GWASs including the genetic diversity in African populations, small sample sizes, allelic heterogeneity of malaria and inherent limitations of the GWAS approaches. In view of providing alternative strategies that can potentially address these challenges, we reviewed the current progress in GWAS and post-GWAS approaches including polygenic analysis, fine-mapping, protein-protein interactions, gene-based and pathway-level analyses. We estimated SNP-heritability of severe malaria at 20.1% in Gambian populations and showed how advanced statistical methods can potentially be implemented in the current malaria resistance datasets to provide useful functional insights. We further advocated for making paradigm shift from single-omics to 'multi-step' and 'multi-dimensional' integrative studies which can integrate multi-layer genetic information from human host and the parasite together with the impacts of environment factors. The future systems biology level studies will eventually revolutionize malaria research and lead to the development of new vaccines and therapeutics.

Damena et al. BMC Medical Genomics (2019) 12:120
https://doi.org/10.1186/s12920-019-0564-x

BMC Medical Genomics

REVIEW Open Access

Genome-wide association studies of severe *P. falciparum* malaria susceptibility: progress, pitfalls and prospects

Delesa Damena[1*], Awany Denis[1], Lemu Golassa[2] and Emile R. Chimusa[1]

Abstract

Background: *P. falciparum* malaria has been recognized as one of the prominent evolutionary selective forces of human genome that led to the emergence of multiple host protective alleles. A comprehensive understanding of the genetic bases of severe malaria susceptibility and resistance can potentially pave ways to the development of new therapeutics and vaccines. Genome-wide association studies (GWASs) have recently been implemented in malaria endemic areas and identified a number of novel association genetic variants. However, there are several open questions around heritability, epistatic interactions, genetic correlations and associated molecular pathways among others. Here, we assess the progress and pitfalls of severe malaria susceptibility GWASs and discuss the biology of the novel variants.

Results: We obtained all severe malaria susceptibility GWASs published thus far and accessed GWAS dataset of Gambian populations from European Phenome Genome Archive (EGA) through the MalariaGen consortium standard data access protocols. We noticed that, while some of the well-known variants including *HbS* and *ABO* blood group were replicated across endemic populations, only few novel variants were convincingly identified and their biological functions remain to be understood. We estimated SNP-heritability of severe malaria at 20.1% in Gambian populations and showed how advanced statistical genetic analytic methods can potentially be implemented in malaria susceptibility studies to provide useful functional insights.

Conclusions: The ultimate goal of malaria susceptibility study is to discover a novel causal biological pathway that provide protections against severe malaria; a fundamental step towards translational medicine such as development of vaccine and new therapeutics. Beyond singe locus analysis, the future direction of malaria susceptibility requires a paradigm shift from single -omics to multi-stage and multi-dimensional integrative functional studies that combines multiple data types from the human host, the parasite, the mosquitoes and the environment. The current biotechnological and statistical advances may eventually lead to the feasibility of systems biology studies and revolutionize malaria research.

Keywords: Genome-wide association study, *P. falciparum* malaria, Susceptibility, Resistance, Heritability, Pathways, Fine-mapping, Multi-omics, Systems biology

* Correspondence: damenadelesa@gmail.com
[1]Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag, Rondebosch, Cape Town 7700, South Africa
Full list of author information is available at the end of the article

Background

Plasmodium falciparum, the causative agent of severe malaria, has been infecting humans for at least 5000-10,000 years following the advent and expansions of agriculture [1–3]. Malaria still poses a huge social, economic and health problems in several low-income countries, particularly in sub Saharan Africa [4, 5]. *P. falciparum* infects millions and kills hundreds of thousands of African children each year. However, this constitutes only a small proportion (1%) of the populations in endemic areas in which the infections progress to severe malaria such as profound anemia or cerebral malaria [6, 7].

Comprehensive understanding of the genetic basis of resistance and susceptibility to severe malaria is crucial to understand the molecular mechanisms of host-parasite interactions that can inform the development of effective therapeutics, vaccination, diagnostics and risk prediction strategies [8, 9]. To this end, GWASs have recently been implemented in malaria endemic areas and replicated some of the well-known variants including *HbS* and *ABO* blood group [10–13]. Despite the fact that malaria is expected to drive several protective alleles to high frequencies that can be captured by GWAS approach, it is unclear why only limited number of novel variants were identified of which a small fraction was replicated across endemic populations. Some of the contributing factors for this discrepancy might include small sample sizes, the genetic diversity of the malaria endemic populations and allelic heterogeneity of malaria protective alleles among others. On the other hand, several association signals distributed across the genome that didn't pass GWAS significance threshold were observed in these studies [10–13]; suggesting the possible existence of polygenic effects. This raises several key questions including 1) What is the genetic architecture of malaria susceptibility/resistance? 2) What is the heritability of malaria susceptibility and its distribution across the genome? and 3) What is the extent and pattern of epistasis and pleiotropy at genome wide scale?

Here we review the current status of malaria susceptibility GWASs and provide guidance to future research directions. We begin by assessing the progress and pitfalls of severe malaria susceptibility GWASs and discuss the biology of the novel variants. We then provide an overview of the recent progresses in post-GWAS approaches and discuss how these methods can be implemented in severe malaria susceptibility studies to better understand the underlying biology. We conclude by discussing on research areas where further works are needed in light of the global malaria eradication efforts.

Results

Severe *P. falciparum* malaria susceptibility GWASs: progress and pitfalls

In malaria endemic areas where repeated *P. falciparum* infection is very common, the majority of children recover from malaria. However, a small proportion of infections progress to the severe form of the disease such as severe anaemia, cerebral malaria, acidosis and respiratory distress [7]. Cerebral malaria is the commonest cause of death characterized by rapid onset of generalized convulsion followed by coma (a Blantyre coma score of less than 3 in the presence of *P. falciparum* parasitaemia). Severe anaemia is defined as a haematocrit of < 15% or haemoglobin < 5 g/dl in the presence of *P. falciparum* parasitaemia [14]. Although the clinical outcome of malaria is determined by several factors including infection rate, parasite genetics and the environment, the host-genetics factor contribute about 25% of *P. falciparum* malaria severity. However only small proportions (~ 2%) of heritability is explained by the well-known variants such as sickle-cell anaemia and α-thalassaemia [7]. The conventional approaches such as candidate gene-based studies [15–17] and the family based linkage studies [18, 19] have been implemented at least for the last three decades and identified several association variants. Unfortunately, the majority of the findings were discordant and failed to replicate in different populations [1].

GWAS in malaria susceptibility study was motivated to address the acute limitations of the conventional approaches and provide better understandings of the underpinning genetics at genome wide scale. To this effect, a global partnership of malaria researchers, named as Malaria Genomic Epidemiology Network (MalariaGEN) was established in 2008 [20]. MalariaGEN has successfully conducted multi-center-scale GWASs [10–13] and reported some interesting findings which we will discuss in later sections. However, the GWAS approach has several limitations including 1) weak performances in genetically diverse populations [8], lack of translation of associated loci into suitable biological hypotheses [21], 3) the well-known problem of missing heritability [22], 4) the lack of understanding of how multiple modestly associated loci within genes interact to influence a phenotype [23], 5) inefficiency in distinguishing between inflation from bias (cryptic relatedness and population stratification) and true signal from polygenicity [24], 6) the imperfection of asymptotic distribution of current mixed model association in the case of low-frequency variants [25]. The discussion on limitations of GWAS approache is beyond the scope of this review. Here we focus on the major challenges of malaria susceptibility GWASs and highlight the recent positive progresses.

Genetic diversity of African population

Owing to the fact that Africa is the origin of modern humans, there is high level of genetic diversity and weak linkage disequilibrium (LD) in Africans compared to non-African populations [26–28]. These distinct genetic characteristics created major setbacks to GWASs in

African populations primarily because of lack of representative dense genotype chips and reference panels [8]. It was estimated that a GWAS of 0.6 million SNPs based on HapMap phase 1 dataset in European population has an equivalent power to the chips with 1.5 million SNPs in African populations [29].

This might have affected the power of previous malaria GWASs. For instance, in the first malaria GWAS [10], *HbS* locus, a well-known variant conferring resistance to severe malaria demonstrated a weak signal (p-values ~ $1x10^{-7}$) because of the weak LD between causal variants and the SNPs that were genotyped. After which authors sequenced the locus, undertook multipoint imputation, used proper reference panel and dramatically improved the signal to p-value ~$1x10^{-14}$. However, coverages of the genotyping chips have been enormously improved to be able to capture the genetic diversities among global populations following the recent technological advances and availability of diverse reference data sets [27, 30–32]. For instance, Omni microarrays based GWASs were proven to have considerable power in African populations [27].

Such developments have also facilitated imputation-based studies in African populations. For instance, Band et al. [9] showed the feasibility of multi-point imputation based meta-analysis in for malaria GWASs using HapMap3 haplotype panel. Another study showed a substantial improvement of imputation accuracy by using the more diversified AGVP WGS reference panel [27]. We believe that the reference dataset will grow further and accelerate genomic research by including wide range of haplotype diversity in African populations.

Sample size

In GWAS, a stringent p-value ($0.5x10^{-8}$) is usually needed to declare evidences of genuine associations to minimize false discovery rate that can arise from multiple testing [33]. Thus, very large sample size is required to achieve genome-wide significance threshold particularly for loci with modest effect sizes. The required sample size is even much higher for studies in population of African ancestry because of the higher genetic diversity. In contrast, the current sample sizes of GWASs in African populations including those of malaria susceptibility are generally small compared to non-Africans [34] which might have affected the power of the studies. Therefore, more powered studies in African population might lead to the discovery of novel association variants.

Allelic heterogeneity of malaria protective variants

Allelic heterogeneity defined as the presence of multiple causal variants in the same locus is one of the challenges of GWA and fine mapping studies [35]. The presence of multiple causal variants with variable effect sizes and LD structures limits the power of GWASs. In such cases, fine-mapping methods will also have lower accuracy to pinpoint true causal variants among several possible candidates [35]. Allelic heterogeneity has been described for the well-known loci affecting malaria susceptibility, which is reflected by their geographical distribution within malaria-endemic regions [36].

Several distinct variants are known to exist at the loci causing inherited hemoglobinopathies [36]. Allele frequencies, LD structure and effect sizes of these variants differ in sub-populations within endemic areas [37]. For instance, the sickle cell allele, *HbS*, is known to have different haplotype structure and effect sizes in different regions of sub-Saharan Africa [27]. *HbC* allele is common in some parts of west Africa such as Burkina Faso, Ghana, Togo and Benin while absent in other west African countries such as Cameroon and Chad [38]. In the same region, several alternative alleles with differing effect sizes are known to exist at the locus causing G6PD deficiency [36, 39]. Although population specific studies can minimize such challenges, the current MalariaGEN datasets are comprised of several populations each with small sample size; making it difficult to undertake powered GWASs for specific geographic areas.

Genetic architecture of malaria susceptibility and resistance

The performance of GWASs is dependent on the genetic architecture of the diseases and traits under investigation. For the majority of complex diseases and traits, the GWAS variants identified thus far, only explain a very small proportion of heritability; a phenomenon commonly termed as 'missing' heritability [22] . There have been different explanations for the 'missing' heritability including common disease rare variant hypothesis [40], none-additive components, primary epistasis [41, 42] and polygenic genetic architecture [21].

One of the challenges of malaria GWASs is that we don't know much about the genetic architecture of malaria protection trait. First, as one of the prominent evolutionary selective forces, the majority of malaria protective alleles might have evolved under positive selection and might potentially be balanced by other forces [43]. In this case, the protective variants are expected to have large effect sizes with high allele frequencies that can be detected by the conventional GWAS approaches; provided that proper reference panel and genotyping platforms are used [43]. Second, similar to the genetic architecture of other infectious diseases [44], malaria protection trait might largely be attributed to few rare variants of large effect sizes. In this case, the GWASs are underpowered as rare variants might not be in LD with common variants. Third, malaria protection trait might be mainly under polygenic and epistatic control [1, 7,

45] which the conventional GWAS approach can't capture.

Biology of the novel variants identified by severe *P. falciparum* malaria GWASs

Severe malaria GWASs have replicated some of the well-known variants such as *HbS* and *ABO* blood groups and few novel variants related to red blood cell membrane biology which reinforce the importance of erythrocyte variants for protection against severe malaria. Besides, the GWASs have identified notable novel association variants in immune and other pathways that may directly or indirectly influence the disease outcome. The epidemiology and biology of the well-known variants were reviewed else-where [1]. Below we characterize the novel malaria susceptibility genetic variants identified by GWASs. We first discuss the biology of two variants such as cluster of the *glycophorin genes (GYPA/B/E)* and *ATP2B4* that were well-replicated across malaria endemic populations. We then extend our discussion to other novel variants.

ATP2B4

The association of variants in *ATP2B4* gene with severe malaria susceptibility was reported by Timmann et al. [11] in Ghanaian populations and replicated in subsequent studies in other populations [12, 46]. SNPs in this locus were also linked with reduction of mean corpuscular hemoglobin concentration (MCHC) level [47]. *ATP2B4* encodes a ubiquitous plasma membrane calcium-transporting protein (*PMCA4b)* [48]. *PMCA4b* is widely expressed in different tissues and is the main transporter of Ca^{2+} in erythrocyte membrane [48].

A recent study showed that the GWAS SNPs are localized in a previously unrecognized *ATP2B4* haplotype named as 'haplotype-1' and individuals with this haplotype exhibit a reduced *PMCA4b* expression level [49]. In this study, it was also shown that the reduction of *PMCA4b* expression significantly decreases the calcium extrusion in RBCs. Consistent with this, a study conducted by Lessard et al. and colleagues elegantly characterized the *ATP2B4* locus using a combination of transcriptomic, epigenomic and gene-editing study approaches [50]. The authors first undertook knock out experiment and demonstrated that *ATP2B4* knocked-out mice express an elevated level of MCHC. Then, they conducted expression quantitative trait locus (eQTL) mapping studies using UK biobank dataset and showed strong associations between *ATP2B4* erythroblast specific variants and RBC related traits including MCHC level, decreased RBC distribution and increased hemoglobin levels.

Further analysis of DNase I hypersensitivity sites (DHSs) at *ATP2B4* and eQTL mapping showed that the GWAS SNPs are mapped to an erythroid specific enhancer element. Deletion of this enhancer from human erythroid cell line using CRISPR-Cas9 system showed a dose dependent reduction of *ATP2B4* expression level. Bi-allelic deletion of the enhancer resulted in eighty three percent reduction of *ATP2B4* expression level compared to the wild type while mono-allelic deletion resulted in moderate reduction of the *ATP2B4* expression level [50].

To determine the effects of the regulatory variants at *ATP2B4* gene on calcium homeostasis, Lessard et al. measured the calcium concentration in unedited and edited (*ATP2B4* enhancer deleted) HUDEP-2 cells. The edited cells demonstrated higher intracellular calcium level compared to wild cells indicating that *ATP2B4* expression is essential for plasma membrane calcium pump. The disturbance of intracellular Ca^{2+} homeostasis might play an important role in impairing the invasion, development and reproduction of malaria parasite in RBCs [51, 52]. Therefore, *ATP2B4* region can potentially be targeted for development of vaccine and therapeutics.

Cluster of the 3 glycophorin genes (GYPA/B/E)

The largest multi-center malaria susceptibility GWAS which included eleven populations was conducted by Band et al. [12]. In this study, 34 genomic regions containing potential susceptibility loci for severe malaria were identified. Among which, a strong signal was observed at locus between *FREM3* gene and cluster of 3 *glycophorin genes (GYPA/B/E*) on Chromosome 4. A haplotype (at SNP *rs184895969)* within this region was reported to reduce the risk of developing severe malaria by about 40% and is common in Kenyan populations with allele frequency reaching 10% [12].

A subsequent study in the same populations identified a large number of copy number variants which are characterized by deletion, duplications and hybrid structures in *GYPA* and *GYPB* genes [53]. Of which a distinct variant called *DUP4* was reported to reduce the risk of severe malaria by about 40% in eastern African (Kenya) populations. Further characterization showed that, this variant is composed of complex *GYPB-A* hybrid and encode *Dantu* antigen in MNS blood group system [53]. The association of this region with severe malaria was supported by another recent case control study in Tanzanian populations [54]. The glycophorin gene cluster, *GYPA* and *GYPB* encode the MNS blood group system and are known to be receptors for *P. falciparum* during RBC invasion [55]. *GYPA* and *GYPB* serve as an erythrocyte membrane receptor for *EBA-175* and *EBL-1* proteins of the parasite respectively [56]. This genomic region is also known to be under an ancient selective

pressure resulted from host-pathogen arm races between *P.falciparum* and humans [57] . Further functional analysis is required to better understand how these variants affect the invasion and/or development of the parasites in erythrocytes and convey protection against severe malaria.

SCO1 and DDC

Notable association signals were identified by the first malaria susceptibility GWAS conducted in Gambian population [10]. The first lead SNP (*rs6503319*) is located close to *SCO1 (synthesis of cytochrome c oxidase)* gene on chromosome 17p13. *SCO1* is a multi-functional signaling protein which plays an essential role in *mitochondrial cytochrome c oxidase (COX)* copper delivery pathways [58]. *COX* catalyzes electron transfer from reduced cytochrome c to oxygen and is abundantly expressed in muscles, brain and liver [58]. Deficiency of *COX* caused by mutations in *SCO1* gene can lead to respiratory distress and severe metabolic acidosis [59] which are also the major complications during cerebral malaria [60]. Further studies are needed to understand how the variants in *SCO1* gene are associated with the pathological pathways of cerebral malaria.

The second notable association signal identified in this study was *Dihydroxypheny-alanine decarboxylase (DDC)* on Chromosome 7p12.2. A recent study in Tanzanian populations replicated the association of *DDC* variants with cerebral malaria [54]. *DDC* gene encodes *Aromatic-L-amino-acid decarboxlase* enzyme which is involved in biosynthesis of neurotransmitters such as dopamine and serotonin [61]. *DDC* is an essential enzyme for brain and nervous developments and its deficiency is associated with reduced cognitive functions [61]. *DDC* is involved in cellular immunity and contributes in protection against parasitic disease in invertebrates [62]. Furthermore, mutations in *DDC* gene was reported to be associated with refractoriness of *Anopheles gambiae* mosquito against *P.falciparum* parasites [63]

MARVELD3

In addition to the *ATP2B4*, Timmann et al. [11] identified an association SNP (*rs2334880*) on chromosome 16p 22.2 which is linked to *MARVELD3*. However, this association has not been replicated in other studies. *MARVELD3* is one of the components of tight junction proteins in several epithelial and endothelial tissues and is expressed as two alternative spliced variants [64]. These proteins are involved in assembly, development, maintenances and regulations of tight junction. Tight junctions play a major role in intracellular adhesions and involved in sub-cellular signaling mechanisms [64].

IL-12 receptors and IL-23 receptors

The most recent malaria susceptibility GWAS was conducted in Tanzanian population [13]. In this study, notable associations signals were identified in immune pathways including in interleukin receptors (*IL-23R and IL-12RBR2*), *in ketch-like proteins (KLHL3)* and Human Leucocyte Antigen (HLA) regions. *Interleukin-12* is formed from a hetrodimer of *IL12B (ILp40 subunit)* and *IL-12A (ILp35 subunit*) [65]. *IL-12* plays a vital role in stimulating cell-mediated immune responses against intra-cellular pathogens through binding to high affinity *IL-12RB1* and *IL-12RBR2* receptor complexes. It promotes the development of *T-helper cells* (*Th1)* and enhances the production of *INF-γ*, both of which are known to mediate the clearance of intracellular pathogens [65]. In malaria, *IL-12* has been implicated in mediating the protective immunity both in experimental animals and in humans [66] . *IL-23* is an important pro-inflammatory cytokine that shares p40 subunits with *IL12* [67]. It induces the differentiation of naive *CD4 T*-cells to *IL-17* which plays key roles in pathogenesis of autoimmune diseases [68].

HLA is encoded by the Major Histocompatibility Complex (MHC), the most polymorphic genes known in human genome. The diversity of MHC is believed to be driven by selection pressure from infectious pathogens and known to be associated with the risk of several infectious diseases [69]. HLA variants such as HLA class I antigen (*HLA-Bw53*) and HLA class II variant (*DRB1*1302-DQB1*0501*) were reported to confer protections against severe malaria in Gambian populations [69]. HLA class I antigen is expressed by liver cells suggesting that T cells (CTL) responses might efficiently act against the liver stage of malaria parasite in individuals with *HLA-Bw53* [69]. On the other hand, individuals with *DRB1*1302-DQB1*0501* variant might possess efficient antigen presentation mechanism that can lead to rapid clearance of blood stage parasites [69].

Variants in immune pathways are of great interest because of their potential to inform the development of effective malaria vaccines [1]. The current study is interesting in that several putative variants in immune pathways were identified. However, the power of this study is limited because of relatively smaller sample size and weak significance threshold used to interpret the findings. Therefore, further studies with higher detective power are needed to consolidate the findings (Table 1).

Polygenic genetic architecture and epistasis: Presenting the absent in the current severe malaria GWASs

Polygenic genetic architecture

Polygenic view of genetic architecture is gaining ground in genetic epidemiological studies and widely implicated for the 'missing' of heritability in GWAS analysis [70] .

Table 1 Summary of the novel severe malaria susceptibility and resistance association variants identified by GWASs

Genomic regions containing the association variants					Genome-wide association studies							
					Jallow *et.al* [10] Pop: Gambian N = 2560 (case =1060, control = 1500)		Timmann et al. [11] Pop: Ghanaian N = 2153 (case =1325, control =828)		Band et al. [12] Pop: African (11 countries) N = 11,552 (case = 5633, control = 5919)		Ravenhall et al. [13] Pop: Tanzanian N = 914 case =449, control = 465)	
Nearest gene name	Chr	Position	SNP ID (Ref/Alt)	MOI	OR	*p*-value	OR	*P*-value	OR	*P*-value	OR	*P*-value
ATP2B4	01	203658471	rs 4951377 (A/G)	DO	–	–	–	–	–	$3.1x10^{-9}$	–	–
		203654024	rs 10900585(T/G)	AD	–	–	0.61	1.9×10^{-10}	–	–	–	–
		203660781	rs4951074(G/A)	AD	–	–	0.62	1.3×10^{-9}	–	–	–	–
IL23R, IL12RB	01	67,731,614	rs6682413(–)	RE	–	–	–	–	–	–	0.48	8×10^{-7}
GYP A/B/E and FREM3	04	143777125	rs184895969(A/C)	DO	–	–	–	–	0.67	9.5×10^{-11}	-	-
C4orf17	04	100429757	rs73832816(–)	REC	–	–	–	–	–	–	0.29	3.8×10^{-7}
AF146191.4–004 (lincRNA)	04	90717704	rs114169033(–)	AD	-	-	-	-	-	-	3.32	6.7×10^{-7}
AC108142.1 (antisense)	04	82822332	rs1878468	HET	-	-	-	-	-	-	0.383	9.0×10^{-7}
Intergenic	05	43,909,343	rs113449872(–)	HET	–	–		–	–	–	0.35	2.2×10^{-8}
KLHL3, MYOT	05	37,011,761	rs2967790(–)	AD	–	–	–	–	–	–	0.60	5.9×10^{-7}
TREML4	06	41,205,690	rs9296359 (–)	HET	–	–	–	–	–	–	4.08	1.2×10^{-7}
DDC	07	50,623,201	rs10249420(C/G)	AD	0.69	6.8×10^{-5}	–	–	–	–	–	–
			rs1451375(–)	DO	0.75	$6.1x10^{-6}$	–	–	–	–	–	–
Intergenic	07	53,676,837	rs17624383(–)	AD	–	–	–	–	–	–	–	5.6×10^{-7}
CSMD1	08	4754838	rs73505850(–)	AD		–	–	–	–	–	4.79	5.9×10^{-7}
LINC00944	12	127237620	rs11335470 (–)	HET	–	–	–	–	–	–	0.40	2.5×10^{-7}
Intergenic	11	130,417,522	rs3133394	AD							0.5	$9.4X10^{-7}$
FAM155A	13	108228013	rs144312179(–)	AD	–	–	–	–	–	–	0.2	6.2×10^{-7}
MARVELD3	16	71,653,637	rs2334880 (T/C)	AD	–	–	1.19	1.9×10^{-6}	–	–	–	–
SOC1	17	10,573,909	rs65033119(–)	AD	1.21	7.2×10^{-7}	–	–	–	–	–	–
Intergenic (LINC00670)	17	12,399,526	rs149085856(–)	AD	–	–	–	–	–	–	3.87	$2.1x10^{-7}$
ZNF536	19	1,069,639	rs8109875(–)	REC	–	–	–	–	–	–	0.5	5.7×10^{-7}

MIO Mode of inheritance, *AD* Additive, *HET* Heterozygous, *DO* Dominant, *REC* Recessive, *OR* Odd-ratio, *Ref* Reference allele, *Alt* Alternative allele, *Pop* Population

The rationale behind polygenic inheritance is that complex-traits/diseases are influenced by multiple variants with modest effects that are too small to pass the stringent genome wide significance threshold [71]. In standard GWAS analysis, 'Genomic control' (GC) method is applied as a quality-control measure to minimize spurious associations that can be caused by population structures such as population stratification and cryptic relatedness.

However, a slight inflation of the test statistics (true but weak signals) which cannot be corrected by GC was initially observed across the genome in a Schizophrenia GWAS [72]. Subsequently, this observation has been supported by other studies [73] and led to the development of a number of statistical tools aiming to capture polygenic signals at genome-wide scale including 1) polygenic scoring method implemented in PLINK software [74] ; 2) Mixed Linear Models (MLM) such as: GCTA [73], BOLT-LMM [75] , Bayes-R [76] and LDAK [77, 78], PCGC [79] 3) linkage-disequilibrium (LD) score regression method [80] among others.

Polygenic contributions in malaria susceptibility and resistance

Co-evolution of host-pathogen model predicts that multiple host loci are involved in resistance/susceptibility to infectious diseases due to the complex interactions between the multi-locus parasite genotype and the corresponding defense from the host-genome [42, 81]. Indeed, malaria might have left multiple genetic variants; the majority of which have effects too small to be detected by the standard GWASs. The existence of polygenic inheritance in malaria protection was predicted by several authors

[1, 7] and supported by the GWASs. For instance, the largest malaria susceptibility GWAS so far, identified 34 regions of the genome containing variants with evidence of associations [12]. Earlier GWAS in Ghanaian population identified 40 genomic regions containing 102 SNPs with evidences of association in the discovery phase of the study [11]. The recent GWAS in Tanzanian populations [13] identified 2322 SNPs at several regions across the genome.

Thus, implementation of polygenic analytic methods in malaria studies may potentially shed more light to the underlying biology. For instance, heritability can be estimated and partitioned in to different cell-types and functional groups and molecular pathways which enable to localize causal variants. Furthermore, these approaches can be extended to explore the genetic correlations between susceptibility to malaria and susceptibility to other infectious diseases. The existence of shared genetic basis between infectious diseases susceptibility/protection is well-documented [82]. However, the extent and pattern of these correlations have not been systematically investigated at genome-wide scale; partly because of inadequate GWAS data for infectious diseases. Such studies can potentially provide clues to common molecular processes between resistance/susceptibility to infectious diseases that will have practical importance including designing multi-purpose vaccine and genetic risk prediction strategies.

Heritability of severe malaria in Gambian population

To figure out how polygenic analysis can be implemented in malaria susceptibility, we accessed the Gambian malaria susceptibility GWAS dataset from European Phenome Genome Archive (EGA) through data application procedure and estimated heritability of malaria susceptibility using MLM approaches. The Gambian GWAS data is the largest MalariaGen dataset obtained from a single country comprised of 4920 samples (2429 cases and 2491controls) and 1.6 million SNPs that passed GWAS quality control (QC). We first excluded the known malaria susceptibility associated loci and performed stepwise extra QC filtering. Specifically, we focused on sample relatedness, SNP missingness proportion and SNP differential missing proportion which are well-known to affect the accuracy of heritability estimation [78]. We then estimated the heritability using GCTA model for different QC thresholds by including 10 principal components (PCs) as fixed effects to account for population structure. As expected, the estimation was unstable when less stringent QC thresholds is applied (varying from 37.8 to 20.1%) as shown in Table 2. However, when more stringent QC (Relatedness threshold (5%), SNP differential missingness proportion ($p \leq 1 \times 10^{-3}$) and SNPs missing proportion of ($p > 0.02$)) was applied, the estimation became stable (~ 20.1%, SE = .05). Neither the inclusion of more PCs (15, 20) as fixed effects nor SNP phasing further brought down the estimate. Using the same stringent QC threshold, the estimation was approximately the same for Mandinka ethnic group (~ 24.3%, SE = 0.6). We couldn't estimate for other ethnic groups because of smaller sample sizes. Furthermore, the use of PCGC model which is designed for case/control approximately showed the same estimate (19.8%, SE = .07).

Although our heritability estimation is fairly stable when stringent QC is implemented, care should be taken in interpreting these results: First, all polygenic methods perform better in less structured data obtained from homogenous populations than the MalariaGen dataset which is comprised of diverse populations spanning most of the Malaria endemic belt in Africa. Second, the methods are designed and perform well in highly polygenic traits/diseases in which effects of each variant is

Table 2 SNP-heritability of severe malaria susceptibility/resistance in Gambian population at different basic quality threshold using MLM

Population	Sample relatedness-threshold	SNP missingness-proportion	SNP differential-missingness Proportion	Prevalence	Covariate principal-components	No. Samples	No. SNPs	GCTA h2(% SE)	PCGC h2(SE)
Gambia	–	5%	–	1%	10	4920	1627656	37.8(.05).	
	5%	5%	1×10^{-10}	1%	10	4128	1627656	30.5(.05)	
	5%	5%	$1x10^{-5}$	1%	10	4128	1607610	28.7(.05)	
	5%	5%	1×10^{-3}	1%	10	4128	1570344	25.1(.05)	
	5%	2%	1×10^{-3}	1%	10	4128	1486554	20.1(.05)	19.8(.07)
	5%	2%	1×10^{-3}	1%	15	4128	1486554	22.5(.05)	
	5%	2%	$1x10^{-3}$	1%	20	4128	1486554	19.5(.05)	
	Phased	–	–	1%	10	4128	1627656	20.4(.06)	
Mandinka	5%	2%	$1x10^{-3}$	1%	10	1281	1486554	24.2(0.6)	

GCTA Genome Complex Trait Analysis, *PCGC* Phenotype Correlation Genotype Correlation regression

mostly modest. In contrast, a considerable proportion of malaria protection trait might be attributed to rare variants of large effect sizes that might not be in LD with common variants and can't not be 'tagged' by SNPs chip which means that the contributions of such variants will not be accounted for. Third, subtle population structure that cannot be corrected by conventional methods such as unmatched case/control can potentially create systematic biases to the estimates.

Epistasis

Epistasis is becoming one of the hot research topics in genetic epidemiological studies in the last few years because of the fact that none additive genetic variations are shown to have significant influence on the phenotype of complex traits/diseases than previously expected [83]. The available statistical approaches and software packages for detection of epistasis at genome-wide scale were reviewed elsewhere [83, 84]. These approaches have been applied in genetic studies of complex diseases such as lupus erythematosus [85], anklosing spondlylitis [86], psiorosis [87] and unraveled previously unknown epistasis interactions between risk loci which explained a significant proportion of 'missing' heritability of the respective diseases.

Epistasis between malaria risk loci have been well documented and implicated as one of the possible reasons for lack of replication of susceptibility variants in different populations and the 'missing' heritability. For instance, sickle-cell trait (*HbS*) and *α thalassaemia* were shown to demonstrate negative epistatic interactions such that the protection against severe malaria offered by *HbS* is reduced when co-inherited with α^{+} *thalassaemia* [88].

A case-control study in Kenyan population also reported that α^{+} *thalassaemia* modulates the effects of Haptoglobin (Hp) variants in predicting the risk of severe malaria [89]. In this study, it was shown that the combination of α^{+} *thalassaemia* and *Hp2–1* variant synergistically increase the protection against severe malaria by about 37%. However, the protective effect is decreased to 13% when α^{+} *thalassaemia* is inherited with *Hp1–1* and further diminished to neutral (zero) when inherited with *Hp2–2*. Similarly, in a multi-center case control study, the existence of negative epistasis interaction between *HbC* and *ATP2B* alleles was reported [15].

Another recent case control study of severe malaria in Kenya reported the existence of negative epistasis between a compliment receptor called *S12* and α^{+}*thalassaemia* in which the protective effect of *S12* higher in children with normal *α-globin* [90]. The extent and pattern of epistatic interaction at genome-wide scale is yet to be explored. In malaria susceptibility GWAS, the priority has been given to a single locus analysis to identify novel risk loci. We expect that, the next step of malaria susceptibility GWAS will include the investigations of epistasis at genome-wide scale.

From GWAS to biology: multi-step and multi-dimensional analyses

Fine-mapping and pathway analyses

The ultimate goal of genetic susceptibility studies is to identify causal variants and understand the underlying biological pathways which can lead to translated medicine such as effective vaccines and therapeutics. However, translating GWAS signals in to biological themes remains an open problem because of the confounding effects from LD between association SNPs, limited knowledge of gene functions and localization of the majority of GWASs hits in none protein coding regulatory regions (regulatory SNPs) [91–93]. In attempt to address this challenge, several fine mapping strategies have been developed and implemented [94]. One such strategy is trans-ethnic fine mapping in which the natural variability of haplotype structure across ethnically diverse populations is used to narrow down candidate causative variants [95]. The smaller LD and diversity of haplotype structure in African population makes it relatively easier to identify the causal SNPs and target gene/genes through fine mapping approaches [8]. However, the fact that malaria protective alleles are heterogeneous across populations might challenge the application of trans-ethnic fine-mapping approaches in malaria susceptibility studies.

Alternatively, several fine mapping statistical tools have recently been developed following the advances in annotation data bases and improved reference panels. These include Bayesian approach, heuristic approach and penalized regression methods [96]. The principles, applications, strength and weakness of these methods is reviewed elsewhere [94, 96]. These methods are increasingly playing crucial role in the efforts being made to pinpoint causal variants of complex diseases/traits. For instance, Galarneau et al. [97] identified novel independent association signals by fine-mapping three loci that are known to influence fetal hemoglobin (HbF) levels. The authors sequenced the three loci (*BCL11A, β-globin and HBS1L-MYB*), undertook dense genotyping, performed step-wise conditional analysis and revealed previously un-recognized SNPs that explain additional genetic variation. Similarly, a recent fine mapping study of HLA region identified several susceptibility loci for multiple infectious diseases [82]. More sophisticated studies that combine statistical and functional fine-mapping strategies have recently been implemented and provided mechanistic insights to the genetic basis of complex diseases [98].

In addition to fine mapping approaches, pathway and interaction analysis can be another avenue for exploring molecular basis of Malaria susceptibility/resistance. Instead of emphasizing on single-variant analysis, these approaches test the coordinated effects of several variants at systems level using biological information from annotation data basis [99]. Pathway analysis improve study power by integrating cumulative effects of weak association signals and provide functional information by identifying associated sets of genes/proteins [100]. By implementing the pathway analysis approaches, several studies have gained new insights in understanding the genetic basis of complex diseases [101–103]. The available statistical tools for pathway and interaction analysis is reviewed in [104]. We therefore, advocate for the implementation of fine mapping and pathway analytic methods in malaria susceptibility studies to shed more light in to the underlying biology.

Epigenomics and Epigenome wide association studies

Epigenetics refers to heritable phenotype changes that do not involve alterations in the DNA sequences such as methylation, post-translational histone modification, histone variation, chromatin remodelling and non-coding RNAs [105]. Epigenetic impacts have recently been implicated in malaria susceptibility and resistance [106, 107]. For instance, in a recent study, strong transcriptional response was detected in monocytes of *P. falciparum* infected individuals from Fulani, an ethnic group that is less susceptible to malaria [107]. The authors suggested that, this response is likely regulated by genome wide chromatin alterations. The discussion on the possible mechanisms of epigenetic impacts on malaria susceptibility and resistance is beyond the scope of this paper and is reviewed in [108].

However, the majorities of epigenetic studies including those of malaria susceptibility and resistance have been limited to either small sample sizes or inadequate genome coverage and thus, lack adequate power to decipher the epigenetic impacts on complex diseases [105]. In effort to address this challenge, investigators recently developed a large-scale, systematic epigenomic equivalents of GWAS called epigenome-wide association studies (EWAS) that attempts to uncover epigenetic variants underlying common diseases/phenotype using genome-wide technologies such as Illumina 450 K array [109]. EWA approach recently gains a considerable attention partly due to the fact that the majorities of the GWAS SNPs are mapped to none coding regions of the genome implying that the variant SNPs cause changes in gene expression levels rather than causing changes in protein function [109]. Thus, combining both genetic (GWA) and epigenetic (EWA) approaches in parallel may prove a fruitful approach for understanding mechanisms of disease risk [109]. Undoubtedly, application of such approaches may shed new light into mechanisms of malaria susceptibility and protection.

Multi-omics approaches

Today, there are significant advances in high throughput technologies that can generate big '-omic' data from all spectrum of molecular biology [91]. The '-omics' studies (Genomics, Epigenomics, Transcriptomics, Proteomics, Metabolomics) are crucial to understand the underpinning biology of complex diseases. In severe malaria, '-omics' studies have provided important clues about the molecular events that lead to either complications or recovery from diseases. For instance, following the discovery of glycophorin regions by the GWAS, a whole genome sequencing-based study [53] was conducted to characterize variants in this region and identified a novel distinct copy number variant called *DUP4* which reduces the risk of severe malaria by about 33% . In addition to this, a genome-wide gene expression study was conducted in Kenyan children and reported increased expression of genes related to neutrophil activation during malaria infections [110]. The authors also observed differential expression of heme- and erythrocytes-related genes in acute malaria patients which reaffirms the importance of erythrocyte-related genes in malaria susceptibility and resistance.

In another host-parasite interaction study, the importance of miRNA in inhibiting parasite growth in erythrocyte was reported [111]. The authors observed translocation of several host RBC miRNAs in to *P. falciparum* parasites, as well as fusion of these human miRNAs with parasite mRNA transcripts to inhibit the translation of enzymes that are vital for the parasite development. Specifically, two micro-RNAs, miRNA-451 and let-7i, were highly enriched in *HbAS* and *HbSS* erythrocytes and these miRNAS along with miR-223 were shown to attenuate the growth of parasite [111].

However, '-omics' studies are limited to single data-type analysis and lack adequate power to explain the complexity of molecular processes and usually lead to identification of correlations than causations [112]. Thus, integrating and analysing multiple '-omics' data enables better understanding of the molecular processes and interactions that give rise to complex diseases/traits. For example, leveraging host microbiome relative abundance data as a second (quantitative) trait, and performing a joint analysis of bivariate phenotypes can increase statistical power by maximizing phenotypic information and inform how the interaction between host genotype with microbiome impacts the phenotype.

Multi-omics approaches aim to integrate big '-omics' data, undertake 'multi-step' and 'multidimensional' analysis for elucidating complex biological problems [112] . Driven by the massive abundance of '-omics' data from

wide ranges of biological molecules, multi-omics strategy have recently provided unprecedented successes in complex diseases/trait studies. The current state of art of multi-omics approach and available statistical methods is recently reviewed in Hasin et al. [112]

Malaria susceptibility and resistance is influenced by several host, parasite and environmental factors as depicted in Fig. 1. The protective alleles have independently evolved in different populations being shaped by the co-evolution and interaction between the human genome, the parasite and the environment [1]. Thus far, single '-omics' data analysis enabled us to understand some of the factors that are associated with the malaria protective traits. To progress beyond associations and pinpoint the causal pathways, it may require to implement carefully designed, coordinated multi-omics studies that involve human host, the parasite, the environment and possibly mosquito. The current advents of high-throughput technologies in generating massive '-omics' data and their continuously decreasing cost complemented with the availability of statistical tools which able to simultaneously capture millions of data points will lead to the implementations of multi-omics approaches in malaria susceptibility studies.

Conclusions and perspectives

The ultimate goal of malaria susceptibility study is to discover a novel causal biological pathway that provide protections against severe malaria; a fundamental step towards translational medicine such as development of vaccine and new therapeutics that can facilitate the global malaria eradication efforts. To achieve this goal, various study approaches have been implemented at least for the last three decades and successfully identified several association variants.

Recently, a number of GWASs have been implemented in malaria endemic areas to better understand the underlying biology. While some of the well-known variants were replicated, only few novel variants were convincingly identified and their biological functions remains to be understood. Several limiting factors including genetic diversity of population in malaria endemic areas, allelic heterogeneity of protective variants, small sample sizes, lack of proper reference panel and

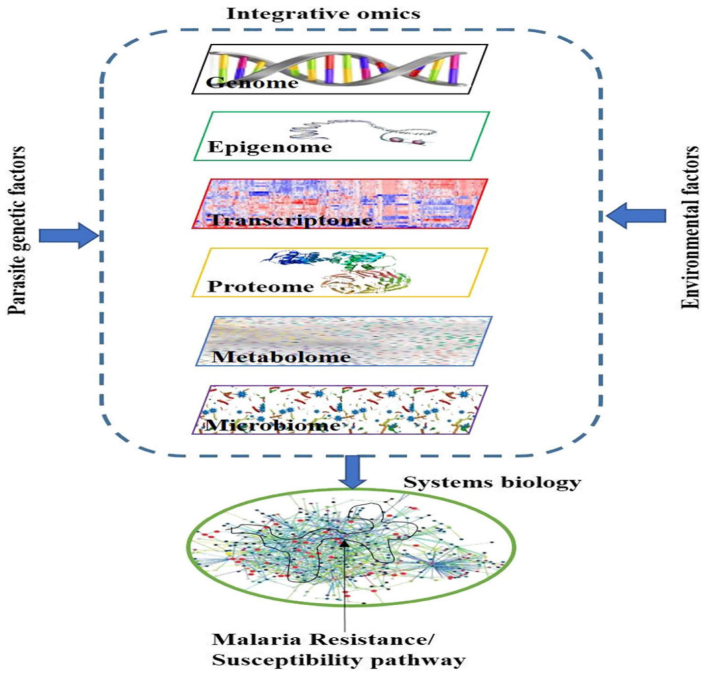


Fig. 1 Schematic representation of the integrative analyses. Systems biology approach which incorporate multiple layers of information from host (multi-omics), the environment and parasite genetic factors can potentially lead to the discovery of malaria protective pathways

proper genotyping chips might have negatively impacted the malaria GWASs.

Another challenge is that we don't know much about the genetic architecture of malaria protective trait. There are at least two scenarios in which GWAS approach might fail; First, malaria protective trait might largely be attributed to rare variants of large effect sizes that might not be in LD with common variants and can't be captured by the GWAS approach. Second, malaria might have left multiple genetic variants distributed across human genome; the majority of which have effects too small to be detected by the standard GWASs [1, 7]. Theoretically, the large sample sizes, dense genotyping chips or whole genome sequencing, use of appropriate reference panels and effective genotype imputation can address the majority of the challenges. However, given the resource constraints; especially, in Africa where malaria problem is the greatest, this will likely take several years to achieve.

On the other hand, the recent advances in statistical techniques is enabling to extract useful information from the present-day GWAS sample sizes. For example, a number of statistical approaches have been developed to capture polygenicity in complex diseases. We showed how these methods can potentially be implemented in malaria susceptibility studies and provide useful insights. We believe that further studies with larger sample sizes can elucidate the polygenic effects in malaria protective trait by extending the analysis to genome partitioning, risk prediction and genetic correlations.

Beyond singe locus analysis, multi-step and multi-locus analyses including pathway analysis, fine mapping and interaction analysis can potentially be implemented in malaria susceptibility GWASs to gain new insights to the underpinning biology. For instance, pathway analysis can provide important information by analyzing the coordinated effects of several variants at systems level using biological information from annotation databases. Methods that combine statistical and functional fine-mapping strategies can potentially be implemented to pinpoint the causal variants from the GWAS association signals.

Most importantly, the future direction of malaria susceptibility requires a paradigm shift from single '-omics' to multi-stage and multi-dimensional integrative functional studies that combines multiple data types from the human host, the parasite, the mosquitoes and the environment. The current biotechnological advances, an ever-increasing annotation data bases and availability of advanced analytical techniques, will eventually lead to feasibility of systems biology studies and revolutionize malaria research.

Abbreviations

eQTL: Expression quantitative trait locus; EWAS: Epigenome wide association studies; GWAS: Genome wide association study; MLM: Mixed Linear Model

Acknowledgements

We thank Kwiatkowski's group from University of Oxford for their constructive comments and assistance. We thank Kirk Rokett and Ellen Leffler for their useful comments. We are very grateful to Gavin Band for his supervision and guidance in heritability analysis and his critical comments. I thank Newton's fund student transfer scheme for funding me during my stay at University of Oxford. We also thank Abdoulaye Djimde for his follow up.

Authors' contributions

DD conceived the work, designed and drafted the manuscript; LG and AD revised the manuscript; EC conceived the work, revised the manuscript and supervised the work. All authors read and approved the final manuscript.

Funding

DD is a PhD student funded by DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, Faculty of Health Sciences, University of Cape Town, Private Bag, Rondebosch, Cape Town 7700, South Africa. [2]Aklilu Lema Institute of Pathobiology, Addis Ababa University, PO box 1176, Addis Ababa, Ethiopia.

Received: 8 December 2018 Accepted: 29 July 2019
Published online: 14 August 2019

References

1. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet. 2005;77:171–92. https://doi.org/10.1086/432519.
2. Parikh S, Rosenthal PJ. Human genetics and malaria : relevance for the design of clinical trials. J Infect Dis. 2008;198:1255–7.
3. Sabeti P, Usen S, Farhadian S, Jallow M, Doherty T, Newport M, et al. CD40L association with protection from severe malaria. Genes Immun. 2002;3:286–91. https://doi.org/10.1038/sj.gene.6363877.
4. Neghina R, Iacobiciu I, Neghina AM, Marincu I. Malaria, a journey in time: in search of the Lost Myths and Forgotten Stories. Am J Med Sci. 2010;340:492–8. https://doi.org/10.1097/MAJ.0b013e3181e7fe6c.
5. World Health Organization. World Malaria Day report. Geneva: WHO; 2017.
6. Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, Marsh V, et al. Indicators of Life-threatening malaria in African children. N Engl J Med. 1995;332:1399–404. https://doi.org/10.1056/NEJM199505253322102.
7. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of malaria in Africa. PLoS Med. 2005:253–1259. https://doi.org/10.1371/journal.pmed.0020340.

8. Teo Y-Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet. 2010;11:149–60. https://doi.org/10.1038/nrg2731.
9. Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, et al. Imputation-Based Meta-Analysis of Severe Malaria in Three African Populations. PLoS Genet. 2013;9. https://doi.org/10.1371/journal.pgen.1003509.
10. Jallow M, Teo YY, Small KS, Rockett KA, Clark TG, Kivinen K, et al. Genmone-wide and fine-resolution association analysis of malria in West Africa. Nat Genet. 2010;41:657–65. https://doi.org/10.1038/ng.388.
11. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012;489. https://doi.org/10.1038/nature11334.
12. Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Band G, Si Le Q, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature. 2015;526:253–7. https://doi.org/10.1038/nature15390.
13. Ravenhall M, Campino S, Sepu N, Nadjm B, Mtove G, Wangai H, et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. PLoS Genet. 2018;14:15. https://doi.org/10.1371/journal.pgen.1007172.
14. World Health Organization. Severe Malaria. Tropical Med and Int Health. 2014;19:7–131. https://doi.org/10.1111/tmi.12313.
15. Malaria Genomic Epidemiology Network. Reappraisal of known malaria resistance loci in a large multicenter study. Nat Genet. 2014;46:1197–204. https://doi.org/10.1038/ng.3107.
16. Verra F, Mangano VD, Modiano D. Genetics of susceptibility to Plasmodium falciparum: From classical malaria resistance genes towards genome-wide association studies. Parasite Immunol. 2009;31:234–53. https://doi.org/10.1111/j.1365-3024.2009.01106.x.
17. Hedrick P. Population genetics of malaria resistance in humans. Heredity (Edinb). 2011;10716:283–304. https://doi.org/10.1038/hdy.2011.16.
18. Flori L, Sawadogo S, Esnault C, Fre N, Fumoux F, Rihet P. Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. Hum Mol Genet. 2003;12:375–8. https://doi.org/10.1093/hmg/ddg033.
19. Rihet P, Traoré Y, Abel L, Aucan C, Traoré-Leroux T, Fumoux F. Malaria in humans: Plasmodium falciparum blood infection levels are linked to chromosome 5q31-q33. Am J Hum Genet. 1998;63:498–505. https://doi.org/10.1086/301967.
20. Achidi EA, Agbenyega T, Allen S, Amodu O, Bojang K, Conway D, et al. A global network for investigating the genomic epidemiology of malaria. Nature. 2008;456:732–7. https://doi.org/10.1038/nature07632.
21. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461:747–53. https://doi.org/10.1038/nature08494.
22. Visscher PM, Wray NR, Zhang Q, Sklar P, Mccarthy MI, Brown MA, et al. 10 Years of GWAS Discovery : biology , function , and translation. Am J Hum Genet. 2017;101:5–22. https://doi.org/10.1016/j.ajhg.2017.06.005.
23. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. Bioinformatics. 2015;32:549–56. https://doi.org/10.1093/bioinformatics/btv619.
24. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed model association methods. Nat Genet. 2014;46:100–6. https://doi.org/10.1038/ng.2876.
25. Daya M, Der Merwe L, Galal U, Möller M, Salie M, Chimusa ER, et al. A panel of ancestry informative markers for the complex five-way admixed South African Coloured population. PLoS One. 2013;8. https://doi.org/10.1371/journal.pone.0082224.
26. Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history, modern human origins, and complex disease mapping. Annu Rev Genomics Hum Genet. 2008;9:403–33. https://doi.org/10.1146/annurev.genom.9.081307.164258.
27. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African genome variation project shapes medical genetics in Africa. Nature. 2015;15(51):327–32. https://doi.org/10.1038/nature13997.
28. Slatkin M. Linkage disequilibrium - understanding the evolutionary past and mapping the medical future. NatRevGenet. 2008;9:477–85. https://doi.org/10.1038/nrg2361.
29. Conrad DF, Jakobsson M, Coop G, Wen X, Wall JD, Rosenberg NA, et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nat Genet. 2006;38:1251–60. https://doi.org/10.1038/ng1911.
30. The International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52–8. https://doi.org/10.1038/nature09298.
31. The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. Nature. 2011;467:1061–73.
32. Johnston HR, Hu YJ, Gao J, O'Connor TD, Abecasis GR, Wojcik GL, et al. Identifying tagging SNPs for African specific genetic variation from the African Diaspora Genome. Sci Rep. 2017;7:1–9. https://doi.org/10.1038/srep46398.
33. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32:381–5. https://doi.org/10.1002/gepi.20303.
34. Popejoy A, Fullerton S. Genomics is failing on diversity. Nature. 2016;538:161–4. https://doi.org/10.1038/538161a.
35. Hormozdiari F, Zhu A, Kichaev G, Ju CJT, Segrè AV, Joo JWJ, et al. Widespread Allelic Heterogeneity in Complex Traits. Am J Hum Genet. 2017;100:789–802.
36. Bauduer F. Red cell polymorphisms and malaria: an evolutionary approach. Bull Mem Soc Anthropol Paris. 2013;25:55–64.
37. Hardison R, David CHK, Belinda G, Cathy R, George P, Patrinos GP, Nicholas A, et al. HbVar: A relational database of human hemoglobin variants and thalassemia mutations at the globin gene server. Hum Mutat. 2002;19:225–33.
38. Piel FB, Howes RE, Patil AP, Nyangiri OA, Gething PW, Bhatt S, et al. The distribution of haemoglobin C and its prevalence in newborns in Africa. Sci Rep. 2013;3:1–8. https://doi.org/10.1038/srep01671.
39. Clark TG, Fry AE, Auburn S, Campino S, Diakite M, Green A, et al. Allelic heterogeneity of G6PD deficiency in West Africa and severe malaria susceptibility. Eur J Hum Genet. 2009;17:1080–5. https://doi.org/10.1038/ejhg.2009.8.
40. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. Rare Allele Hypotheses for complex diseases. Curr Opin Genet Dev. 2010;19:212–9. https://doi.org/10.1016/j.gde.2009.04.010.
41. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci. 2012;109:1193–8. https://doi.org/10.1073/pnas.1119675109.
42. Hall MD, Ebert D. The genetics of infectious disease susceptibility: Has the evidence for epistasis been overestimated? BMC Biol. 2013;11. https://doi.org/10.1186/1741-7007-11-79.
43. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15:379–93. https://doi.org/10.1038/nrg3734.
44. Wilfert L, Schmid-Hempel P. The genetic architecture of susceptibility to parasites. BMC Evol Biol. 2008;8:1–8. https://doi.org/10.1186/1471-2148-8-187.
45. Mitchell-Olds T, Willis JH, Goldstein DB. Which evolutionary processes influence natural genetic variation for phenotypic traits? Nat Rev Genet. 2007;8:845–56. https://doi.org/10.1038/nrg2207.
46. Bedu-Addo G, Meese S, Mockenhaupt FP. An ATP2B4 polymorphism protects against malaria in pregnancy. J Infect Dis. 2013;207:1600–3. https://doi.org/10.1093/infdis/jit070.
47. Li J, Glessner JT, Zhang H, Hou C, Wei Z, Bradfield JP, et al. GWAS of blood cell traits identifies novel associated loci and epistatic interactions in caucasian and African-American children. Hum Mol Genet. 2013;22:1457–64. https://doi.org/10.1093/hmg/dds534.
48. Guerini D, Pan B, Carafoli E. Expression, purification, and characterization of isoform 1 of the plasma membrane Ca2+ pump. Focus on calpain sensitivity. J Biol Chem. 2003;278:38141–8. https://doi.org/10.1074/jbc.M302400200.
49. Zámbó B, Várady G, Padányi R, Szabó E, Németh A, Langó T, et al. Decreased calcium pump expression in human erythrocytes is connected to a minor haplotype in the ATP2B4 gene. Cell Calcium. 2017;65:73–9. https://doi.org/10.1016/j.ceca.2017.02.001.
50. Lessard S, Gatof ES, Beaudoin M, Schupp PG, Sher F, Ali A, et al. An erythroid-specific ATP2B4 enhancer mediates red blood cell hydration and malaria susceptibility. J Clin Invest. 2017;127:3065–74. https://doi.org/10.1172/JCI94378.
51. Tiffert T, Lew VL, Ginsburg H, Krugliak M, Croisille L, Mohandas N. The hydration state of human red blood cells and their susceptibility to invasion by Plasmodium falciparum. Blood. 2005;105:4853–60. https://doi.org/10.1182/blood-2004-12-4948.
52. Gazarini ML, Thomas AP, Pozzan T, Garcia CRS. Calcium signaling in a low calcium environment: How the intracellular malaria parasite solves the problem. J Cell Biol. 2003;161:103–10. https://doi.org/10.1083/jcb.200212130.

53. Leffler EM, Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, et al. Resistance to malaria through structural variation of red blood cell invasion receptors. Science. 2017;356:1140–52. https://doi.org/10.1126/science.aam6393.
54. Manjurano A, Sepulveda N, Nadjm B, Mtove G, Wangai H, Maxwell C, et al. USP38, FREM3, SDC1, DDC, and LOC727982 Gene Polymorphisms and Differential Susceptibility to Severe Malaria in Tanzania. J Infect Dis. 2015; 212:1129–39. https://doi.org/10.1093/infdis/jiv192.
55. Ko WY, Kaercher KA, Giombini E, Marcatili P, Froment A, Ibrahim M, et al. Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. Am J Hum Genet. 2011;88:741–54. https://doi.org/10.1016/j.ajhg.2011.05.005.
56. Mayer DCG, Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, et al. Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. Proc Natl Acad Sci. 2009;106:5348–52. https://doi.org/10.1073/pnas.0900878106.
57. Wang HY, Tang H, Shen CKJ, Wu CI. Rapidly evolving genes in human. I. The glycophorins and their possible role in evading malaria parasites. Mol Biol Evol. 2003;20:1795–804. https://doi.org/10.1093/molbev/msg185.
58. Leary SC. Redox regulation of SCO protein function: controlling copper at a mitochondrial crossroad. Antioxid Redox Signal. 2010;13:1403–16. https://doi.org/10.1089/ars.2010.3116.
59. Valnot I, Osmond S, Gigarel N, Mehaye B, Amiel J, Cormier-Daire V, et al. Mutations of the SCO1 gene in mitochondrial cytochrome c oxidase deficiency with neonatal-onset hepatic failure and encephalopathy. Am J Hum Genet. 2000;67:1104–9. https://doi.org/10.1016/S0002-9297(07)62940-1.
60. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature. 2002;415:673–9. https://doi.org/10.1038/415673a.
61. Martelle SE, Raffield LM, Palmer ND, Cox AJ, Freedman BI, Hugenschmidt CE, et al. Dopamine pathway gene variants may modulate cognitive performance in the DHS - Mind Study. Brain Behav. 2016;6:1–12.
62. Zhou Z, Yang J, Wang L, Zhang H, Gao Y, Shi X, et al. A dopa decarboxylase modulating the immune response of scallop chlamys farreri. PLoS One. 2011;6:e18596.
63. Menge DM, Zhong D, Guda T, Gouagna L, Githure J, Beier J, et al. Quantitative trait loci controlling refractoriness to Plasmodium falciparum in natural Anopheles gambiae mosquitoes from a malaria-endemic region in western Kenya. Genetics. 2006;173:235–41.
64. Steed E, Rodrigues NTL, Balda MS, Matter K. Identification of MarvelD3 as a tight junction-associated transmembrane protein of the occludin family. BMC Cell Biol. 2009;10:1–14.
65. Huang D, Cancilla MR, Morahan G. Complete primary structure, chromosomal localisation, and definition of polymorphisms of the gene encoding the human interleukin-12 p40 subunit. Genes Immun. 2000;1:515–20.
66. Luty AJF, Perkins DJ, Lell B, Schmidt-Ott R, Lehman LG, Luckner D, et al. Low interleukin-12 activity in severe Plasmodium falciparum malaria. Infect Immun. 2000;68:3909–15.
67. Duvallet E, Semerano L, Assier E, Falgarone G, Boissier MC. Interleukin-23: A key cytokine in inflammatory diseases. Ann Med. 2011;43:503–11.
68. Minke B, Wu C-F, Pak W. Genes required for cytotoxicity against virus-infected target cells in *K* and *D* regions of H-2 complex. Nature. 1975; 254:84–7.
69. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Benett S, Brewster D, McMichael AJ, Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991;15:595–600.
70. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016;40:268–72.
71. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
72. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460: 748–52.
73. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for whole-genome association and population-based linkage Analyses. Am J Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.
75. Loh P, Tucker G, Bulik-sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts; 2015. https://doi.org/10.1038/ng.3190.
76. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Erratum to "Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels" (J. Dairy Sci. 95:4114–4129). J Dairy Sci. 2014;97:6622. https://doi.org/10.3168/jds.2014-97-10-6622.
77. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91:1011–21. https://doi.org/10.1016/j.ajhg.2012.10.010.
78. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. Nat Genet. 2017;49:986–92.
79. Golan D, Lander ES, Rosset S. Measuring missing heritability: Inferring the contribution of common variants. Proc Natl Acad Sci. 2014;111:E5272–81. https://doi.org/10.1073/pnas.1419064111.
80. Bulik-sullivan BK, Loh P, Finucane HK, Ripke S, Yang J, Working S, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:1–7. https://doi.org/10.1038/ng.3211.
81. Frank SA. Recognition and Polymorphism in host-parasite genetics. Philos Trans Biol Sci. 1994;346:283–93 https://doi.org/10.1098/rstb.1994.0145.
82. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun. 2017;8. https://doi.org/10.1038/s41467-017-00257-5.
83. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nature Rev Genet. 2014;15:722–33. https://doi.org/10.1038/nrg3747.
84. Cordell HG. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009;10:392–404.
85. Castillejo-López C, Delgado-Vega AM, Wojcik J, Kozyrev SV, Thavathiru E, Wu YY, et al. Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. Ann Rheum Dis. 2012;71:136–42.
86. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011;43:761–7.
87. Strange A, Capon F, Spencer CCA, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction betwEn HLA-C and ERAP1. Nat Genet. 2010;42:985–90.
88. Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S, et al. Negative epistasis between the malaria-protective effects of α + − thalassemia and the sickle cell trait. Nature Genet. 2005;37:1253–7.
89. Atkinson SH, Uyoga SM, Nyatichi E, Macharia AW, Nyutu G, Ndila C, et al. Epistasis between the haptoglobin common variant and α +thalassemia influences risk of severe malaria in Kenyan children. Blood. 2014;123:2008–16.
90. Opi DH, Swann O, Macharia A, Uyoga S, Band G, Ndila CM, et al. Two complement receptor one alleles have opposing associations with cerebral malaria and interact with α+thalassaemia. Elife. 2018;7:1–27.
91. van der Sijde MR, Ng A, Fu J. Systems genetics: From GWAS to disease pathways. Biochim Biophys Acta - Mol Basis Dis. 1842;2014:1903–9. https://doi.org/10.1016/j.bbadis.2014.04.025.
92. Boyle EA, Li YI, Pritchard JK. Leading Edge Perspective An Expanded View of Complex Traits. Cell. 2017;169:1177–86.
93. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common. Science (80-). 2012;337:1190–5.
94. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24:111–9.
95. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet. 2011; 43:55–9. https://doi.org/10.1038/ng.732.
96. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018;19:491–504.
97. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet. 2010;42:1049–51.
98. Farh KKH, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518:337–43.
99. Vaughan AM, Aly AS, Kappe SHI. Malaria parasite pre-erythrocytic stage infection: Gliding and Hiding. Cell Host Microbe. 2008;4:209–18.

53. Leffler EM, Band G, Busby GBJ, Kivinen K, Le QS, Clarke GM, et al. Resistance to malaria through structural variation of red blood cell invasion receptors. Science. 2017;356:1140–52. https://doi.org/10.1126/science.aam6393.
54. Manjurano A, Sepulveda N, Nadjm B, Mtove G, Wangai H, Maxwell C, et al. USP38, FREM3, SDC1, DDC, and LOC727982 Gene Polymorphisms and Differential Susceptibility to Severe Malaria in Tanzania. J Infect Dis. 2015; 212:1129–39. https://doi.org/10.1093/infdis/jiv192.
55. Ko WY, Kaercher KA, Giombini E, Marcatili P, Froment A, Ibrahim M, et al. Effects of natural selection and gene conversion on the evolution of human glycophorins coding for MNS blood polymorphisms in malaria-endemic African populations. Am J Hum Genet. 2011;88:741–54. https://doi.org/10.1016/j.ajhg.2011.05.005.
56. Mayer DCG, Cofie J, Jiang L, Hartl DL, Tracy E, Kabat J, et al. Glycophorin B is the erythrocyte receptor of Plasmodium falciparum erythrocyte-binding ligand, EBL-1. Proc Natl Acad Sci. 2009;106:5348–52. https://doi.org/10.1073/pnas.0900878106.
57. Wang HY, Tang H, Shen CKJ, Wu CI. Rapidly evolving genes in human. I. The glycophorins and their possible role in evading malaria parasites. Mol Biol Evol. 2003;20:1795–804. https://doi.org/10.1093/molbev/msg185.
58. Leary SC. Redox regulation of SCO protein function: controlling copper at a mitochondrial crossroad. Antioxid Redox Signal. 2010;13:1403–16. https://doi.org/10.1089/ars.2010.3116.
59. Valnot I, Osmond S, Gigarel N, Mehaye B, Amiel J, Cormier-Daire V, et al. Mutations of the SCO1 gene in mitochondrial cytochrome c oxidase deficiency with neonatal-onset hepatic failure and encephalopathy. Am J Hum Genet. 2000;67:1104–9. https://doi.org/10.1016/S0002-9297(07)62940-1.
60. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature. 2002;415:673–9. https://doi.org/10.1038/415673a.
61. Martelle SE, Raffield LM, Palmer ND, Cox AJ, Freedman BI, Hugenschmidt CE, et al. Dopamine pathway gene variants may modulate cognitive performance in the DHS - Mind Study. Brain Behav. 2016;6:1–12.
62. Zhou Z, Yang J, Wang L, Zhang H, Gao Y, Shi X, et al. A dopa decarboxylase modulating the immune response of scallop chlamys farreri. PLoS One. 2011;6:e18596.
63. Menge DM, Zhong D, Guda T, Gouagna L, Githure J, Beier J, et al. Quantitative trait loci controlling refractoriness to Plasmodium falciparum in natural Anopheles gambiae mosquitoes from a malaria-endemic region in western Kenya. Genetics. 2006;173:235–41.
64. Steed E, Rodrigues NTL, Balda MS, Matter K. Identification of MarvelD3 as a tight junction-associated transmembrane protein of the occludin family. BMC Cell Biol. 2009;10:1–14.
65. Huang D, Cancilla MR, Morahan G. Complete primary structure, chromosomal localisation, and definition of polymorphisms of the gene encoding the human interleukin-12 p40 subunit. Genes Immun. 2000;1:515–20.
66. Luty AJF, Perkins DJ, Lell B, Schmidt-Ott R, Lehman LG, Luckner D, et al. Low interleukin-12 activity in severe Plasmodium falciparum malaria. Infect Immun. 2000;68:3909–15.
67. Duvallet E, Semerano L, Assier E, Falgarone G, Boissier MC. Interleukin-23: A key cytokine in inflammatory diseases. Ann Med. 2011;43:503–11.
68. Minke B, Wu C-F, Pak W. Genes required for cytotoxicity against virus-infected target cells in *K* and *D* regions of H-2 complex. Nature. 1975; 254:84–7.
69. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Benett S, Brewster D, McMichael AJ, Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. Nature. 1991;15:595–600.
70. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016;40:268–72.
71. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565–9.
72. International Schizophrenia Consortium. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature. 2009;460: 748–52.
73. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82. https://doi.org/10.1016/j.ajhg.2010.11.011.
74. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for whole-genome association and population-based linkage Analyses. Am J Hum Genet. 2007;81:559–75. https://doi.org/10.1086/519795.
75. Loh P, Tucker G, Bulik-sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts; 2015. https://doi.org/10.1038/ng.3190.
76. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Erratum to "Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels" (J. Dairy Sci. 95:4114–4129). J Dairy Sci. 2014;97:6622. https://doi.org/10.3168/jds.2014-97-10-6622.
77. Speed D, Hemani G, Johnson MR, Balding DJ. Improved heritability estimation from genome-wide SNPs. Am J Hum Genet. 2012;91:1011–21. https://doi.org/10.1016/j.ajhg.2012.10.010.
78. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ. Reevaluation of SNP heritability in complex human traits. Nat Genet. 2017;49:986–92.
79. Golan D, Lander ES, Rosset S. Measuring missing heritability: Inferring the contribution of common variants. Proc Natl Acad Sci. 2014;111:E5272–81. https://doi.org/10.1073/pnas.1419064111.
80. Bulik-sullivan BK, Loh P, Finucane HK, Ripke S, Yang J, Working S, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47:1–7. https://doi.org/10.1038/ng.3211.
81. Frank SA. Recognition and Polymorphism in host-parasite genetics. Philos Trans Biol Sci. 1994;346:283–93 https://doi.org/10.1098/rstb.1994.0145.
82. Tian C, Hromatka BS, Kiefer AK, Eriksson N, Noble SM, Tung JY, et al. Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. Nat Commun. 2017;8. https://doi.org/10.1038/s41467-017-00257-5.
83. Wei W-H, Hemani G, Haley CS. Detecting epistasis in human complex traits. Nature Rev Genet. 2014;15:722–33. https://doi.org/10.1038/nrg3747.
84. Cordell HG. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet. 2009;10:392–404.
85. Castillejo-López C, Delgado-Vega AM, Wojcik J, Kozyrev SV, Thavathiru E, Wu YY, et al. Genetic and physical interaction of the B-cell systemic lupus erythematosus-associated genes BANK1 and BLK. Ann Rheum Dis. 2012;71:136–42.
86. Evans DM, Spencer CCA, Pointon JJ, Su Z, Harvey D, Kochan G, et al. Interaction between ERAP1 and HLA-B27 in ankylosing spondylitis implicates peptide handling in the mechanism for HLA-B27 in disease susceptibility. Nat Genet. 2011;43:761–7.
87. Strange A, Capon F, Spencer CCA, Knight J, Weale ME, Allen MH, et al. A genome-wide association study identifies new psoriasis susceptibility loci and an interaction betwEn HLA-C and ERAP1. Nat Genet. 2010;42:985–90.
88. Williams TN, Mwangi TW, Wambua S, Peto TEA, Weatherall DJ, Gupta S, et al. Negative epistasis between the malaria-protective effects of α + − thalassemia and the sickle cell trait. Nature Genet. 2005;37:1253–7.
89. Atkinson SH, Uyoga SM, Nyatichi E, Macharia AW, Nyutu G, Ndila C, et al. Epistasis between the haptoglobin common variant and α +thalassemia influences risk of severe malaria in Kenyan children. Blood. 2014;123:2008–16.
90. Opi DH, Swann O, Macharia A, Uyoga S, Band G, Ndila CM, et al. Two complement receptor one alleles have opposing associations with cerebral malaria and interact with α+thalassaemia. Elife. 2018;7:1–27.
91. van der Sijde MR, Ng A, Fu J. Systems genetics: From GWAS to disease pathways. Biochim Biophys Acta - Mol Basis Dis. 1842;2014:1903–9. https://doi.org/10.1016/j.bbadis.2014.04.025.
92. Boyle EA, Li YI, Pritchard JK. Leading Edge Perspective An Expanded View of Complex Traits. Cell. 2017;169:1177–86.
93. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common. Science (80-). 2012;337:1190–5.
94. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24:111–9.
95. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Shi Y, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet. 2011; 43:55–9. https://doi.org/10.1038/ng.732.
96. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018;19:491–504.
97. Galarneau G, Palmer CD, Sankaran VG, Orkin SH, Hirschhorn JN, Lettre G. Fine-mapping at three loci known to affect fetal hemoglobin levels explains additional genetic variation. Nat Genet. 2010;42:1049–51.
98. Farh KKH, Marson A, Zhu J, Kleinewietfeld M, Housley WJ, Beik S, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015;518:337–43.
99. Vaughan AM, Aly AS, Kappe SHI. Malaria parasite pre-erythrocytic stage infection: Gliding and Hiding. Cell Host Microbe. 2008;4:209–18.

100. Wang L, Zhang B, Wolfinger RD, Chen X. An integrated approach for the analysis of biological pathways using mixed models. PLoS Genet. 2008;4:e1000115.
101. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009;10:191–201.
102. Liu C, Bousman CA, Pantelis C, Skafidas E, Zhang D, Yue W, et al. Pathway-wide association study identifies five shared pathways associated with schizophrenia in three ancestral distinct populations. Transl Psychiatry. 2017;7:e1037.
103. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris DW, et al. Molecular pathways involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia and bipolar disorder susceptibility. Mol Psychiatry. 2011;16:286–92.
104. Kao PYP, Leung KH, Chan LWC, Yip SP, Yap MKH. Pathway analysis of complex diseases for GWAS, extending to consider rare variants, multi-omics and interactions. Biochim Biophys Acta - Gen Subj. 1861;2017:335–53.
105. Xu J, Zhao L, Liu D, Hu S, Song X, Li J, et al. EWAS: epigenome-wide association study software 2.0. Bioinformatics. 2018;34:2657–8.
106. Gupta H, Chaudhari S, Rai A, Bhat S, Sahu PK, Hande MH, et al. Genetic and epigenetic changes in host ABCB1 influences malaria susceptibility to Plasmodium falciparum. PLoS One. 2017;12:e0175702.
107. Quin JE, Bujila I, Chérif M, Sanou GS, Qu Y, Homann MV, et al. Major transcriptional changes observed in the Fulani, an ethnic group less susceptible to malaria. Elife. 2017;6:1–19.
108. Arama C, Quin JE, Kouriba B, Farrants AKÖ, Troye-Blomberg M, Doumbo OK. Epigenetics and malaria susceptibility/protection: A missing piece of the puzzle. Front Immunol. 2018;9:1–8.
109. James F. Epigenome-Wide Association Studies (EWAS): Past, Present, and Future 2015. Methods Mol Biol. 2015;1238:51–63.
110. Griffiths MJ, Shafi MJ, Popper SJ, Hemingway CA, Kortok MM, Wathen A, et al. Genomewide analysis of the host response to malaria in Kenyan children. J Infect Dis. 2005;191:1599–611.
111. Lamonte G, Philip N, Reardon J, Lacsina JR, Chapman L, Thornburg CD, et al. Translocation of sickle cell erythrocyte microRNAs intoPlasmodium falciparum inhibits parasite translation and contributes to malaria resistance. Cell Host Microbe. 2012;12(2):187–99. https://doi.org/10.1016/j.chom.2012.06.007.
112. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. Genome Biol. 2017;18:1–15.

Publisher's Note

3. CHAPTER THREE: GENOME-WIDE HERITABILITY ANALYSIS OF SEVERE MALARIA RESISTANCE REVEALS EVIDENCE OF POLYGENIC INHERITANCE

Summary

In this chapter, we investigated the polygenic effects of severe malaria trait in three African populations including Gambia, Kenya and Malawi. We estimated SNP-heritability (h^2g) of severe malaria resistance from genotype GWAS dataset and determined its distribution across the genome including allele frequency spectrum, chromosomes and genomic annotations. In addition to the direct estimation from the raw genotype, we performed various functional enrichment analysis from GWAS-summary statistics meta-analysed across the three populations using stratified linkage disequilibrium score regression (LDSC) method. The LDSC approach require population specific reference panel to produce reliable results making it difficult to apply in African populations which have been poorly represented in public reference domain.

In effort to address this challenge, we prepared African population specific reference panel and computed the LD scores. By using LD information from the prepared panel, we estimated the univariate h^2g and partitioned it in to cell-specific and nonspecific functional categories. We estimated the h^2g estimated at 0.21 (se = 0.05, P = 2.7×10^{-5}), 0.20 (se = 0.05, P = 7.5×10^{-5}) and 0.17 (se = 0.05, P = 7.2×10^{-4}) in Gambian, Kenyan and Malawi populations, respectively. A comparable range of h^2g [0.21 (se = 0.02, P < 1×10^{-5})] was estimated from GWAS-summary statistics meta-analysed across the three populations. We showed for the that the h^2g of malaria resistance is largely ascribed by common SNPs and the causal variants are overrepresented in protein coding regions of the genome. Our African population specific reference panel can potentially be used in other LD-based analyses involving African populations. Our current SNP-heritability estimates were roughly close to a report from a previous family-based study. This might be due to the fact that the previous study underestimated the heritability estimates:

Human Molecular Genetics, 2020, Vol. 29, No. 1 168–176

doi: 10.1093/hmg/ddz258
Advance Access Publication Date: 6 November 2019
Bioinformatics Article

BIOINFORMATICS ARTICLE

Genome-wide heritability analysis of severe malaria resistance reveals evidence of polygenic inheritance

Delesa Damena and Emile R. Chimusa*

Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine University of Cape Town, Private Bag, Rondebosch, 7700 Cape Town, South Africa.

***To whom correspondence should be addressed at.** Division of Human Genetics, Department of Pathology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa. Tel: +27 21 406 6425;Email: emile.chimusa@.uct.ac.za. http://web.cbio.uct.ac.za/~emile/index.html

Abstract

Background: Estimating single nucleotide polymorphism (SNP)-heritability (h^2_g) of severe malaria resistance and its distribution across the genome might shed new light in to the underlying biology. Method: We investigated h^2_g of severe malaria resistance from a genome-wide association study (GWAS) dataset (sample size = 11 657). We estimated the h^2_g and partitioned in to chromosomes, allele frequencies and annotations using the genetic relationship-matrix restricted maximum likelihood approach. We further examined non-cell type-specific and cell type-specific enrichments from GWAS-summary statistics. Results: The h^2_g of severe malaria resistance was estimated at 0.21 (se = 0.05, $P = 2.7 \times 10^{-5}$), 0.20 (se = 0.05, $P = 7.5 \times 10^{-5}$) and 0.17 (se = 0.05, $P = 7.2 \times 10^{-4}$) in Gambian, Kenyan and Malawi populations, respectively. A comparable range of h^2_g [0.21 (se = 0.02, $P < 1 \times 10^{-5}$)] was estimated from GWAS-summary statistics meta-analysed across the three populations. Partitioning analysis from raw genotype data showed significant enrichment of h^2_g in genic SNPs while summary statistics analysis suggests evidences of enrichment in multiple categories. Supporting the polygenic inheritance, the h^2_g of severe malaria resistance is distributed across the chromosomes and allelic frequency spectrum. However, the h^2_g is disproportionately concentrated on three chromosomes (chr 5, 11 and 20), suggesting cost-effectiveness of targeting these chromosomes in future malaria genomic sequencing studies. Conclusion: We report for the first time that the heritability of malaria resistance is largely ascribed by common SNPs and the causal variants are overrepresented in protein coding regions of the genome. Further studies with larger sample sizes are needed to better understand the underpinning genetics of severe malaria resistance.

Introduction

In spite of the global eradication efforts, malaria remains a major global public health problem with 219 million cases and 435 000 deaths in 2017 (1). *Plasmodium falciparum* malaria results in diverse clinical manifestations ranging from asymptomatic parasitaemia to severe malaria (2). Such a wide variation of clinical outcome is attributed to several factors including genetic factors of the host, virulence of parasite and environmental factors (2). Family studies reported that the host genetic factors (heritability) contributes ∼25% of the variations observed in clinical severity of malaria in endemic populations (3). Thus, understanding the molecular basis of the natural immunity against severe malaria will speed up the development of an efficient malaria vaccine.

Driven by the wide availability of genome-wide single nucleotide polymorphism (SNP) arrays, the focus of genetic studies has been shifted from the traditional candidate gene and family-based linkage studies to GWASs. However, only a small fraction of narrow sense heritability is explained by the GWAS

Received: May 17, 2019. **Revised:** October 14, 2019. **Accepted:** October 23, 2019

significant SNPs. This led to a problem commonly termed as 'missing heritability' (4). In effort to address this problem, several statistical methods that aim at quantifying SNP-heritability without identifying the causal variants were developed (5–7).

Packages such as the genome-wide complex trait analysis (GCTA) implement genetic relationship-matrix restricted maximum likelihood (GREML) method in which all SNPs of unrelated subjects are simultaneously analysed in a linear mixed model framework to estimate the proportion of phenotypic variations explained by genotypic variation (5). Apart from GREML approaches, Golan *et al.* (7) recently introduced a regression method called phenotype-correlation-genotype-correlation (PCGC) for estimation of heritability from case-control datasets. PCGC is a Haseman–Elston regression model in which a normalized phenotype is regressed on the genetic covariances of all unique pair of samples. The slope of the regression is used as an estimator of h^2_g. Application of these methods provided new insights in to the genetic architecture of complex diseases including autism, schizophrenia, Parkinson's disease, type 2 diabetes, hypertension among others (8–13). Alternative contemporary statistical methods that enable estimation of h^2_g from publicly available GWAS-summary statistics without the need of individual genotype data are also widely available (14–16) and gained popularity due to their privacy advantages and computational costs.

Because only unrelated individuals are included, the h^2_g analysis offers a greater flexibility of study designs that enable us to conduct powered studies. The use of only unrelated individuals also minimizes the biases from shared environments, one of the greatest challenges in pedigree studies (5). Furthermore, h^2_g analytic methods allow partitioning of the cumulative heritability in to different functional categories and biological pathways and thus, provide more insights in to the underpinning biology (6,17).

Even though a number of severe malaria GWASs have recently been implemented in endemic areas in Africa and reported few novel association variants (18–21), little is known about the h^2_g and its distribution across the genome. Here we present results from a comprehensive h^2_g study of severe malaria resistance in three African populations including Gambia, Kenya and Malawi. We estimated h^2_g and partitioned in to chromosomes, different minor allele frequency (MAF) bins, functional categories and cell types. We found that the h^2_g is disproportionately concentrated on three chromosomes (chr 5, 11 and 20) and enriched in the coding region of genome. Overall, our results suggest that malaria resistance is mainly under polygenic control.

Results

SNP-heritability estimates from genotype datasets

We estimated h^2_g at different quality control (QC) levels to determine the appropriate threshold (see Materials and Methods). As expected, the estimates were inflated at less stringent QC thresholds (Supplementary Material, Fig. S1). However, applying stringent QC protocols including relatedness threshold (5%), SNP differential missingness proportion ($P \leq 1 \times 10^{-3}$) and SNPs missing proportion ($P > 0.02$) yielded a more stable ranges of h^2_g values that were not affected by the inclusion of additional principal components (15, 20, 50) as covariates. At the stringent QC threshold, the h^2_g of severe malaria resistance was 0.21 (se = 0.05, $P = 2.7 \times 10^{-5}$), 0.20 (se = 0.05, $P = 7.5 \times 10^{-5}$) and 0.17 (se = 0.05, $P = 7.2 \times 10^{-4}$) in Gambian, Kenyan and Malawi populations, respectively (Table 1). These estimates were approximately similar in Kenya ethnic groups such as Chonye (0.20, se = 0.07, $P = 5.1 \times 10^{-3}$) and Giriama (0.19, se = 0.07, $P = 7.3 \times 10^{-3}$). However, the estimate was slightly inflated in Mandinka ethnic group of Gambia (0.24, se = 0.06, $P = 5.1 \times 10^{-5}$). We did not estimate h^2_g for other ethnic groups because of inadequate sample sizes. Furthermore, the PCGC model yielded broadly similar results including 0.20, (se = 0.06, $P = 9.7 \times 10^{-4}$), 0.16 (se = 0.06, $P = 8 \times 10^{-3}$) and 0.23 (se = 0.07, $P = 1.3 \times 10^{-3}$) in Gambia, Kenya and Malawi populations, respectively.

Proportion of SNP-heritability attributable to GWAS loci. To quantify the effects of the known variants, we estimated the h^2_g without removing the severe malaria GWAS loci from the datasets (see Materials and Methods). This resulted in slight increment of the h^2_g estimate in Gambian [0.27 (se = 0.05, $P = 1 \times 10^{-5}$)] and Kenyan [0.26 (se = 0.05, $P < 1 \times 10^{-5}$)] populations. Repeating the GREML analysis by including *rs334* as an additional covariate decreased the estimate to [0.24 (se = 0.05, $P < 1 \times 10^{-5}$)] and [0.23 (se = 5%, $P < 1 \times 10^{-5}$)] in Gambian and Kenyan populations, respectively. This suggests that the h^2_g attributable to the GWAS significant loci and *HbS* locus is approximately 0.07 and 0.03, respectively.

Partitioning SNP-heritability by chromosomes, minor allele frequencies and functional annotations. We observed no significant differences in h^2_g estimate obtained from separate analysis (0.24, se = 0.05, $P < 1 \times 10^{-5}$) and the joint analysis (0.22, se = 0.05, $P = 1 \times 10^{-5}$) (Supplementary Material, Fig. S2), suggesting that the population structure was adequately controlled. Moreover, we observed significant correlations between chromosomal length and h^2_g per chromosome (Adj $r^2 = 0.38$, $P = 0.001$) (Fig. 1). However, the estimates of three chromosomes (chr 5, 11 and 20) and three other chromosomes (chr 7, 8 and 15) fell above and below the expected h^2_g at 95% CI, respectively. Notably, chr5 contained a considerable proportion (~0.035) of the h^2_g (Supplementary Material, Table S1).

We performed MAF stratified analysis to estimate the relative contribution of variants with various allele frequencies (Fig. 2). However, we did not find significant differences between the proportion of h^2_g attributed to different MAF bins [standards errors overlapped at 95% confidence interval (CI)]. Moreover, the total sum of our h^2_g estimate per bin [0.27 (se = 0.08, $P = 5.3 \times 10^{-5}$)] was not significantly different from the univariate estimate. We further estimated the h^2_g explained by genic SNPs and the intergenic SNPs at 0.165 (se = 0.05) and 0.062 (se = 0.05), respectively (Table 2). On average, a SNP residing in genic region was enriched 2.9× compared to a SNP residing in an intergenic region of the genome. This is statistically significant at 95% CI.

Functional enrichment from GWAS-summary statistics

After imputation of severe malaria GWAS-summary statistics and QC filtering, we obtained a total of 20 million high quality SNPs (see Materials and Methods). Using this dataset, we estimated the liability scale h^2_g at 0.21 (se = 0.02, $P < 1 \times 10^{-5}$). Partitioning the h^2_g in to 24 main genomic annotations (baseline model) showed evidences of enrichment in multiple categories including 5′UTR (11×), digital genomic footprint (DGF; 10×), enhancer (9×), coding (6×), H3K4me1 (4.9×), TSS (5×), transcription factor binding sites (TFBS; 4×) and FANTOM enhancer

Table 1. h^2_g of severe malaria resistance determined by GCTA and PCGC methods

Population	SNPs (*n*)	Samples (*n*)	$h^2_{g\text{-GCTA}}$ (%)	$h^2_{g\text{-PCGC}}$ (%)
Gambia	1 513 822	4128	0.20(se = 0.05)	0.20(se = 0.05)
Kenya	1 579 227	2062	0.20(se = 0.05)	0.16 (se = 0.05)
Malawi	1 502 462	2418	0.17(se = 0.05)	0.23(se = 0.06)
Mandinka	1 513 822	1281	0.24 (se = 0.06)	ne*
Chonye	1 579 227	637	0.20(se = 0.06)	ne*
Giriama	1 579 227	1173	0.19 (se = 0.06)	ne*

ne*: Model did not fit because of small sample size and there was no reliable estimation SNP: single nucleotide polymorphisms, $h^2_{g\text{-GCTA}}$: h^2_g estimated using GCTA method, $h^2_{g\ PCGC}$: h^2_g estimated using PCGC method

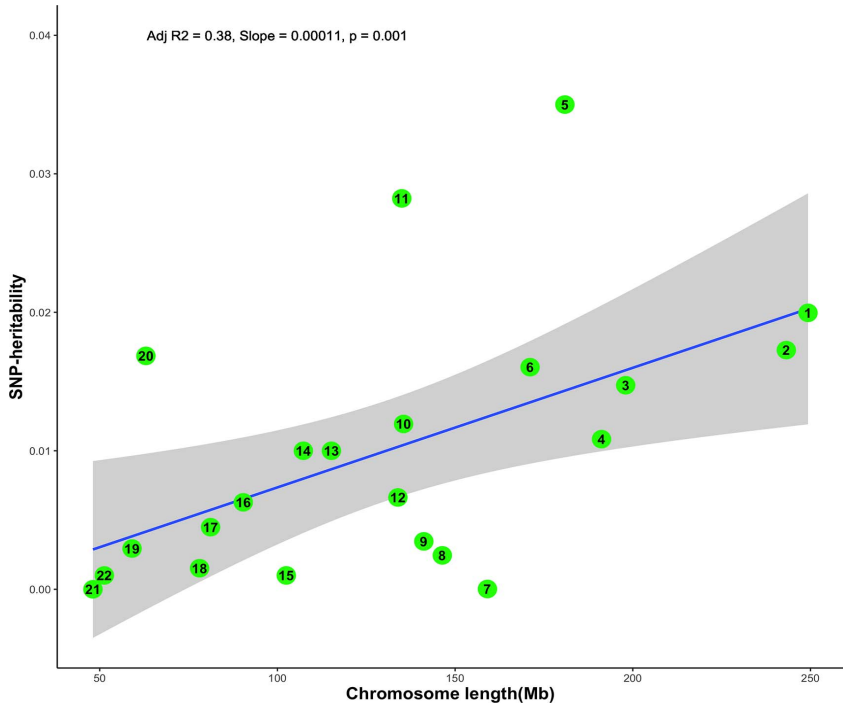


Figure 1. h^2_g per chromosome(y-axis) plotted against chromosome length (x-axis). The blue line represents the h^2_g estimates regressed against chromosome length. The grey shaded areas represent the 95% CI around the slope of the regression model.

Table 2. h^2_g of severe malaria resistance partitioned in to genic and intergenic genomic regions

SNP location	SNPs(*n*)	10 kb boundary h^2_g	h^2_g per SNP
Genic	727 996	0.165(se = 0.05)	2.3×10^{-5}
Inter-genic	785 826	0.062(se = 0.05)	7.9×10^{-6}

(4×) as shown in Figure 3. However, none of the enrichments was statistically significant after correction for multiple testing. Further cell-type specific and cell group analysis did not show significant enrichments.

Discussions

In this study, we estimated the h^2_g and functional enrichment of malaria resistance in three African populations and their meta-analysis. After excluding the severe malaria resistance GWAS loci, we performed GREML analysis at different QC levels to determine the appropriate threshold; indeed, the estimates were inflated upward at less stringent QC levels and became stable at more stringent QC levels. These estimates were broadly similar across the three study populations. Except a slight inflation observed in Mandinka ethnic group which might have been underpowered because of small sample size, the estimates were also similar across the major

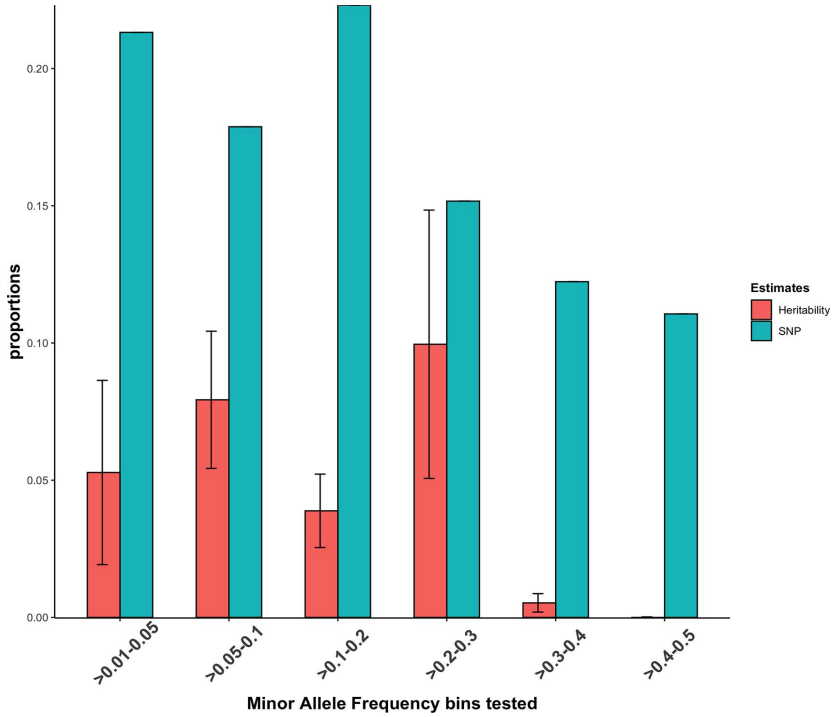


Figure 2. h^2_g partitioned in to different allele frequency spectrum. We created six MAF-bins and estimated the proportion of h^2_g attributed to each bin. The proportion of h^2_g attributed to each bin is shown in red bar and the proportion of SNPs per bin is shown by the blue bar. Error bars represent the 95% CI of the estimate.

ethnic groups. Approximately a similar range of h^2_g of severe malaria resistance was recently reported (22). This might suggest that substantial human genetic factors that influence malaria disease severity have been maintained across endemic populations. Consistent with our findings, a previous family-based study reported a similar range of heritability of severe malaria resistance in two different endemic populations in Kenya (4).

In contrast to the findings from other complex disease studies in which h^2_g is much smaller than family-based heritability values (23), our current h^2_g estimates were roughly close to a report from a previous family-based study (4). This might be due to the fact that the previous study underestimated the heritability estimates: First, only additive genetic effects (narrow-sense heritability) was calculated i.e. the contributions of nonadditive effects including epistasis, dominance and gene-gene interactions were not taken in to account. Second, the authors indicated that their paternity assessment was prone to misclassification, which might have underestimated the actual narrow-heritability estimate (4).

In the current study, including GWAS significant SNPs in the GREML analysis resulted in an increment of the h^2_g estimates by $\sim$0.07 in the study populations, suggesting that the h^2_g attributable to the known malaria resistance GWAS loci ($h^2_{g\text{-GWAS}}$) is generally small. This is consistent with the hypothesis that the vast proportion of heritability of complex traits/diseases is explained by SNPs with effect sizes too small to attain the stringent genome wide significance threshold ($P < 5 \times 10^{-8}$) at the current sample sizes (24). Repeating the analysis by including *rs334* as an additional covariate brought down these estimates by $\sim$0.03, suggesting that more than a third of the $h^2_{g\text{-GWAS}}$ is attributable to the *HbS* locus. This might be explained by the fact that the *HbS* locus has relatively larger effect sizes in the endemic populations (18).

To gain better insights in to the genetic architecture of severe malaria resistance, we partitioned the h^2_g in to different chromosomes, allele frequency spectrum and annotations. Separate GREML analysis and joint analysis yielded broadly similar h^2_g estimates, suggesting that population structure is adequately controlled. The rationale is that. However, the joint analysis in which genomic relatedness matrix (GRMs) of all chromosomes are simultaneously fitted in to a single GREML model, can effectively control the upward biases that can be created by correlated SNPs on different chromosomes (29).

Supporting the polygenic view of genetic architecture, we found a correlation between h^2_g per chromosome and chromosomal lengths (Adj $r^2 = 0.39$, $P = 0.001$). However, the h^2_g is disproportionately concentrated on three chromosomes (chr 5, 11 and 20), suggesting that these chromosomes might contain loci with larger effects against the polygenic background. Thus,

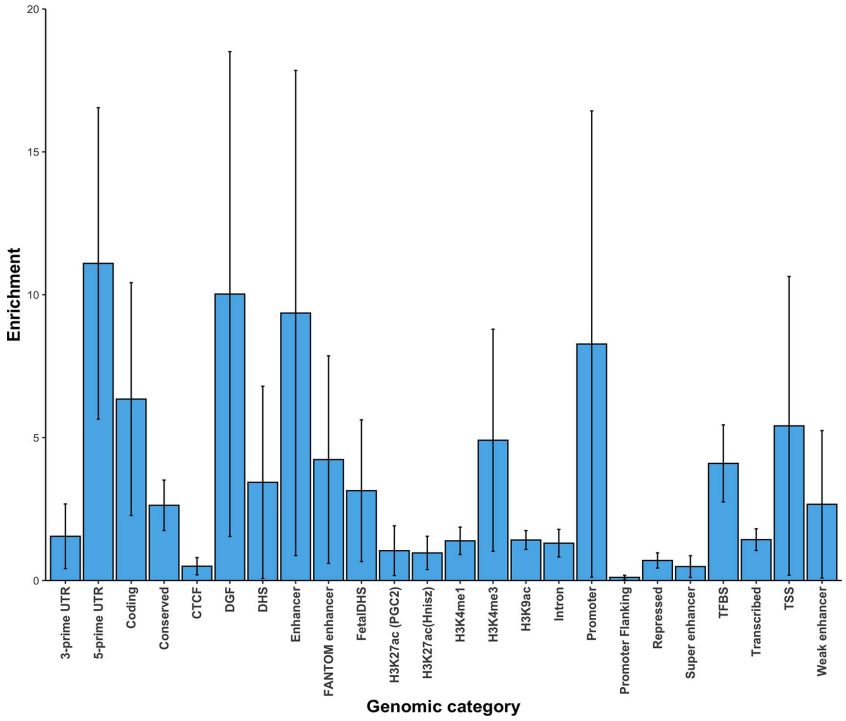


Figure 3. Enrichment estimates of h^2_g for the 24 main annotations. Error bars represent jackknife standard errors around the estimates of enrichment.

targeting these chromosomes using more powered studies (e.g. DNA sequencing) might be a cost-effective approach to discover new severe malaria resistance loci. Previous family-based studies reported that a region on chr5 (5q31–q33) is associated with susceptibility to mild malaria (25,26).

MAF-stratified analysis didn't reveal significant differences between the proportion of h^2_g attributed to different MAF bins. This might assert that h^2_g of severe malaria resistance is broadly uniform across the allele frequency spectrum and is not over-represented by rare alleles. Partitioning by annotation revealed that the h^2_g of severe malaria resistance is significantly enriched in SNPs residing in protein coding regions of the genome, suggesting that further studies focusing on coding regions (e.g. exome sequencing and/or exome array genotyping) might lead to the discovery novel variants.

In addition to the direct estimation of h^2_g from raw genotype datasets, we performed functional enrichment analysis from GWAS-summary statistics using stratified linkage disequilibrium score regression (LDSC) approach (14). To improve the performance of the analysis, we created a reference panel that is more specific to our study populations by merging the African population datasets obtained from 1000 Genomes Project and African Genome Variation Project (27). Using this panel, we created annotation files and estimated h^2_g of severe malaria resistance from GWAS-summary statistics meta-analysed across the study populations.

Our liability scale h^2_g estimate [0.21 (se = 0.02, $P < 1 \times 10^{-5}$)] was comparable to the direct estimates from raw genotype datasets. However, our functional enrichment analyses did not reveal significant results. One of the downsides of stratified LDSC method is that it requires very large sample sizes to detect significant enrichments (14). Of note, the coding genes and the surrounding categories were among the top annotations in our base line model. This further highlights the importance of protein coding regions of the genome in influencing the malaria disease severity.

Finally, our study had a number of caveats that might directly or indirectly affect the accuracy of estimating the true genetic heritability. First, the controls used in this study were not screened for mild malaria that might potentially bias the accuracy of h^2_g estimates. Second, assumptions of the models implemented for the analyses might not adequately explain the true genetic architecture of severe malaria resistance. Third, all the models implemented here do not measure the variances attributable to environmental factors. Fourth, the study is underpowered for the functional enrichment anlyses.

Conclusions

In conclusion, our study showed for the first time that heritability of severe malaria resistance is largely explained by common SNPs and is disproportionately enriched in SNPs residing

in protein coding regions of the genome. Consistent with the polygenic genetic architecture, we observed that the h^2_g of severe malaria resistance is distributed across chromosomes and allele frequency spectrum. However, the h^2_g is disproportionately concentrated on three chromosomes (chr 5, 11 and 20), suggesting the cost-effectiveness of targeting these chromosomes in future malaria genomic sequencing studies. In this study, we created annotation files using population specific reference panel and showed that stratified LDSC analysis can provide reliable SNP-heritability estimates in African populations. Further studies with larger sample sizes are needed to understand the unpinning genetics and biology of severe malaria resistance trait.

Materials and Methods

Description of the study datasets

GWAS datasets of three African populations including Gambia, Kenya and Malawi were obtained from European Phenome Genome Archive (EGA) through the MalariaGen consortium standard data access protocols (28,29). The datasets contain information about a total of 11 657 samples including 4921 samples from Gambia (2491 cases and 2430 controls), 3752 samples from Malawi (3752 cases and 3220 controls) and 2984 samples from Kenya (1506 cases and 1478 controls). Cases were obtained from children who were admitted to Hospitals and fulfilled WHO case definition for severe malaria (29) and controls were obtained from the general population (18–21). All the samples were genotyped on Illumina Omni 2.5 M array. Information about phenotypes, imputation and QC was also provided.

Quality control

The basic QC protocols including plate effects, sample relatedness, Hardy–Weinberg equilibrium, heterozygosity, missingness and differential missingness were done as described elsewhere (18,30). Taking in to consideration that small artifacts can have substantial cumulative effects in h^2_g analysis (31), we applied further stringent QC filtering steps. Briefly, we aligned the quality filtered VCF files to forward stand of the human reference sequence (GRCh3) using the illumina supplied files (www.well.ox.ac.uk/~wrayner/strand) and removed all SNPs with position and strand mismatches. We further removed SNPs with MAF below 0.01, deviate from Hardy–Weinberg at *P*-value below 0.01 using PLINK software (32). We then implemented step-wise QC filtering based on SNPs missingness proportion, differential missingness and sample relatedness as described in (6).

Estimating heritability from genotype data

We applied GCTA (5) and PCGC (7) models to estimate the h^2_g of severe malaria resistance from raw genotype datasets. Briefly, we excluded the region of extended inversion (7 238 552–12 442 658) on chromosome 8p23 (33), the major histocompatibility complex (MHC) region (25 000 000–40,000,000) on chr 6 and the known severe malaria resistance loci including the *ATP2B4* region on chr1:203 154 024–204 154 024, cluster of *glycophorin* (*GYPA/B/E*) region on chr4:143 000 000–146 000 000, *ABO* blood group region on chr9:135 630 000–136 630 000, and the sickle cell (*HbS*) region on chr11:2 500 000–6 500 000 to avoid potential biases from large effects.

We constructed GRMs from pruned high quality independent autosomal SNPs using GCTA software (5) and obtained list of samples with relatedness threshold >5%. We then computed GRMs using all the autosomal SNPs for each cohort and excluded one of any pair of samples with relatedness threshold >5% as recommended elsewhere (6). The final sample of unrelated individuals was 4128, 2062 and 2418 for Gambia, Kenya and Malawi, respectively. The distribution of off-diagonal element of the GRMs for each population is shown in Supplementary Material, Fig. S3.

We used population prevalence of 1% of severe malaria as previously described in (29) and included the top 10 PCs as fixed effects in the GREML analysis. We then transformed the estimates to liability scale as described in Lee *et al.* (34). Using the same GRMs, we estimated the h^2_g using PCGC model as outlined in Golan *et al.* (12). We also computed separate GRMs and estimated h^2_g for major ethnic groups in Gambia (Mandinka) and Kenya (Girimia and Chonye). Furthermore, we created GRMs in the presence of the GWAS significant SNPs and performed GREML analysis to quantify the effects of malaria resistance GWAS loci. We repeated the analysis by including *rs334* as additional covariate to estimate the h^2_g attributable to *HbS*.

Partitioning SNP-heritability from genotype data. Using Gambian dataset (largest sample size), we partitioned h^2_g by chromosomes, MAF bins and annotations. For the partitioning analyses, we excluded the severe malaria resistance GWAS loci to minimize the potential biases from SNPs with large effects. To investigate the biases that might be created by population structure, we performed separate and joint GREML analysis using all autosomal chromosomes. We first computed GRMs for individual autosomal chromosome and estimated h^2_g attributed to each chromosome by separate GREML in which one chromosome is fitted to the model at a time. We then performed a joint analysis in which GRMs of all autosomal chromosomes are simultaneously fitted in to a single GREML analysis and compared the results obtained from both analyses.

In addition to this, we partitioned the h^2_g in to different allele frequencies and annotations. Briefly, we created five MAF bins including > 0.01–0.05, > 0.05–0.1, > 0.1–0.2, > 0.2–0.3, > 0.3–04, > 0.4–0.5, computed separate GRMs for each bin and performed joint GREML analysis. For partitioning by annotation, we mapped all the autosomal SNPs to the human reference panel hg19 in UCSC genome database (http://genome.ucsc.edu) using QCTOOLV2 (https://www.well.ox.ac.uk/~gav/qctool) and obtained a list of genic and intergenic variants. Genic variants included those SNPs mapped to genomic regions within 10 kilobases (kb) upstream and downstream of a protein coding gene. Intergenic variants included all the SNPs mapped to genomic regions outside 10 kb of a protein coding gene. We constructed separate GRMs and estimated h^2_g attributable to each category using the joint analysis implemented in GCTA software (5).

Functional enrichment analysis of SNP-heritability from GWAS-summary statistics

African-specific reference panel. Partitioning h^2_g in to cell-types and functional categories using stratified LDSC approach has recently been shedding new lights in to the genetic architecture of several complex diseases (14,35,36). The method is based on the fact that a given category of SNPs is enriched for h^2_g if SNPs with high LD to that category have higher χ^2 statistics than SNPs with low LD to that category (35,36). However, the

stratified LDSC analysis require population specific reference panel and very large sample sizes to produce reliable results (14). Consequently, the current European 1000G haplotype reference panel that is used as a default in LDSC software (14,35) might not well represent our study populations.

To address this challenge, we created a reference panel that matches with our study populations. Briefly, we merged African population datasets obtained from 1000 Genomes Project and African Genome Variation Project (27) based on overlapped variants and removed structural variants and ambiguous SNPs using plink tool (32). This resulted in a combined dataset of sample size ($n = 4975$). After excluding the admixed populations including Americans of African Ancestry and African Caribbean, we clustered the dataset in to East African and west African sub-regions using smart pca software (37) as shown in Supplementary Material, Fig. S4. We removed SNPs with MAF $< 1\%$, missingness > 0.05 and HWE in controls (alpha level 0.0001), and retained a total of 22 473 268 SNPs (sample size = 2112) and 18 919 068 SNPs (sample size = 380) in east African and west African sub-regions, respectively. We finally calculated the MAF of the panel for later partitioning analysis. Owing to the fact that our study populations are comprised of both east African (Malawi and Kenya) and west Africa (Gambian) populations, we used the entire dataset as a reference panel for functional enrichment analysis.

Baseline model and functional annotations. We created baseline model and cell type specific annotations for our reference panel as described in (14). The baseline-LD model included 24 main annotations that are not cell type-specific including coding, UTRs (3′UTR and 5′UTR), promoter and intronic regions obtained from UCSC genome browser and processed by Gusev *et al.* (12), the histone marks (H3) such as: acetylation of histone at lysine 9 (H3K9ac), monomethylation (H3K4me1) and trimethylation (H3K4me3) of H3 at lysine 4 obtained from Trynka *et al.* (38), acetylation of H3 at lysine 27(H3K27ac) version one processed by Hnisz *et al.* (39) and version two Psychiatric Genomics Consortium, combined chromHMM and Segway predictions obtained from Hoffman *et al.* (40), regions that are conserved in mammals (41,42), super enhancers (39), FANTOM5 enhancers (43), TFBS and DGF post-processed by Gusev *et al* (12). Around each partition, we added 500 bp windows as separate categories to prevent biases that might arise from adjacent annotations.

The 24 main annotations together with the additional windows and a category containing all SNPs yielded 53 overlapping baseline model. Next, we created 220 cell type-specific annotations for the four histone marks: H3K4me1, H3K4me3, H3K9ac and H3K27ac (14) using our reference panel and computed LD score for each annotation. We then combined the 120 cell specific annotations in to 10 cell groups including adrenal and pancreas, central nervous system, cardiovascular, connective and bone, gastrointestinal, immune and hematopoietic, kidney, liver, skeletal muscle and other as described in (14). For each of the 10 categories, we computed the corresponding LD scores.

Stratified LDSC analysis. We obtained meta-analysed GWAS-summary statistics of the three populations ($n = 15\,122\,094$ SNPs) from the previous GWAS (18). We performed imputation on this dataset using ImpG software (44). Briefly, we removed SNPs that mismatch with 1000G phase three markers, computed z-score from the association statistics and performed the imputation using ImGv.1.1 under default settings. We used all the 661 individuals labeled as 'AFRICAN' haplotypes in phase 1 of 1000 Genome Project version-3 calls (45). We removed all imputed SNPs with a predicted accuracy less than 0.9 and SNPs with MAF < 0.01. After QC filtering, we performed stratified LD score regression analysis using our reference panels as described in (14). Briefly, we converted the summary statistics to LDSC format, filtered SNPs with imputation accuracy greater than nine and MAF $> 1\%$, removed structural variants, ambiguous SNPs, the MHC region and significant SNPs. We then performed non-cell type- and cell type- specific analyses as described in (14).

Funding

National Research Foundation of South Africa for Funding (NRF) (grant RA171111285157/119056); The Centre for High-Performance Computing (CHPC, www.chpc.ac.za) for computing resources; The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) program (grant PD00217ML to D.D.); NIH projects (to E.C.).

Conflict of Interest statement. The authors declare no competing interests.

Authors Contributions

DD designed, performed the data analysis and drafted the manuscript, EC contributed in designing, data-analysis and revision of the manuscript and supervised the work.

Acknowledgements

We thank Kwiatkowski's group from University of Oxford for their constructive comments and assistance. We thank Gavin Band for his advice and guidance in heritability analysis from genotype data. This work was supported through the DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

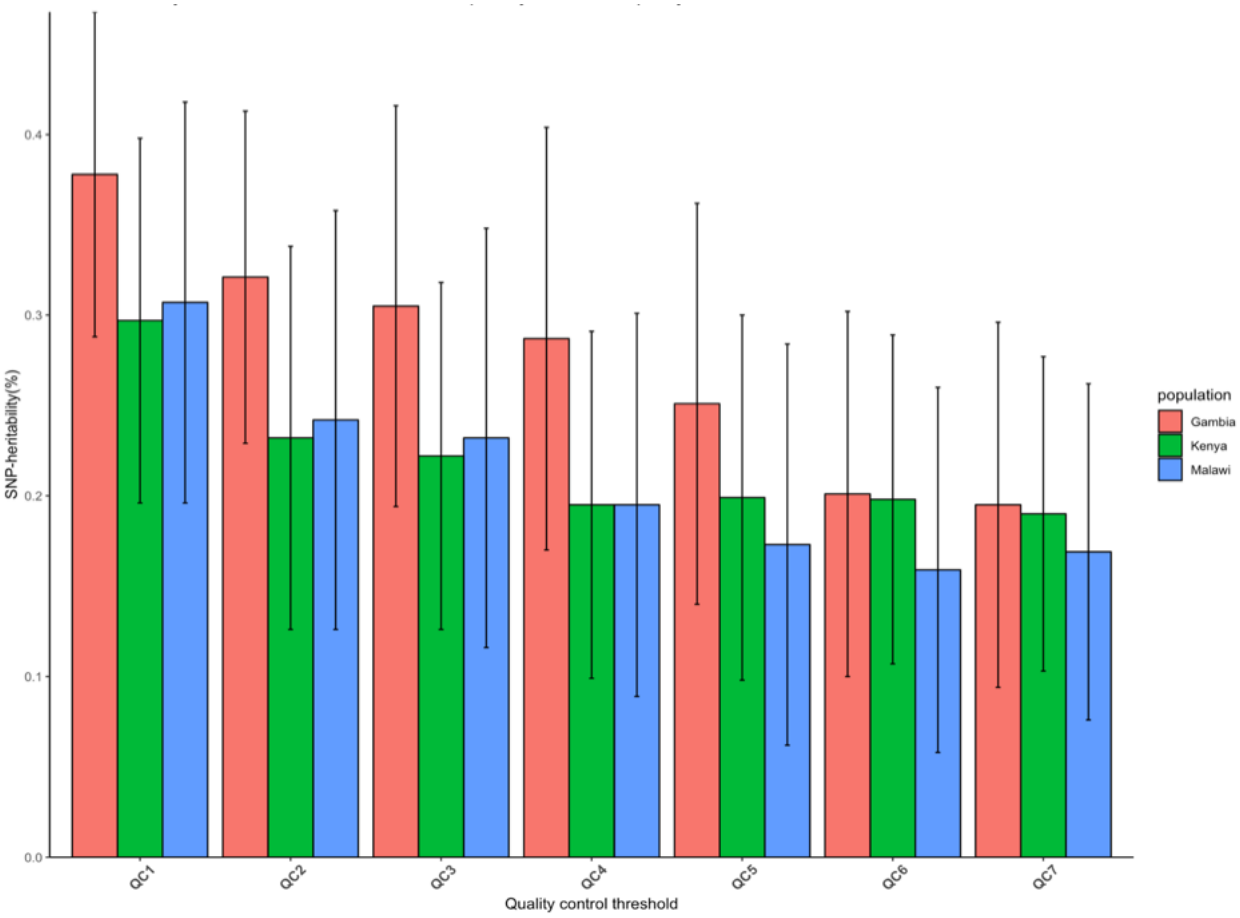
References

1. World Health Organization (2018) *World malaria report*. World Health Organization, https://apps.who.int/iris/handle/10665/275867.
2. Kwiatkowski, D.P. (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *Am. J. Hum. Genet.*, **77**, 171–192.
3. Mackinnon, M.J., Mwangi, T.W., Snow, R.W., Marsh, K. and Williams, T.N. (2005) Heritability of malaria in Africa. *PLoS Med.*, **12**, e340 10.1371/journal.pmed.0020340.
4. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., Mccarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A. *et al.* (2009) Finding the missing heritability of complex diseases. *Nature*, **461**, 747–753.

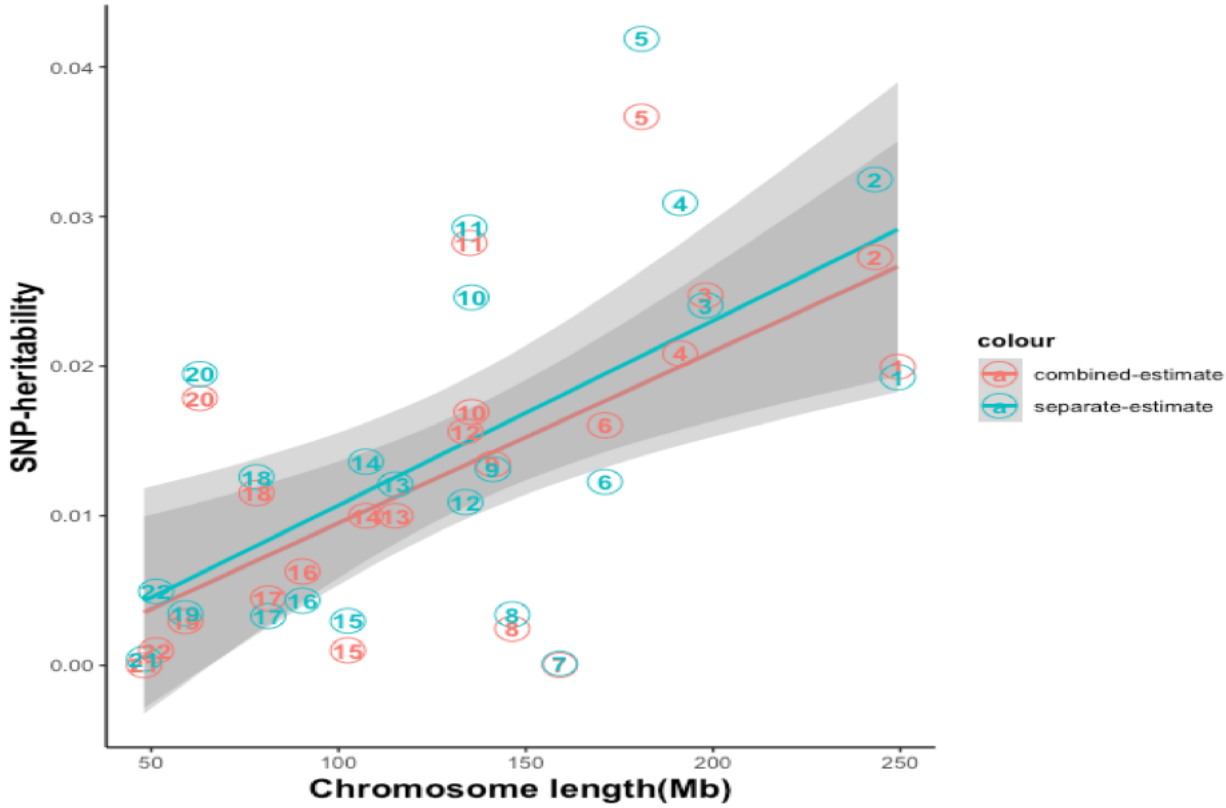
5. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
6. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S. and Balding, D.J. (2017) Reevaluation of SNP heritability in complex human traits. *Nat. Genet.*, **49**, 986–992.
7. Golan, D., Lander, E.S. and Rosset, S. (2014) Measuring missing heritability: inferring the contribution of common variants. *Proc. Natl. Acad. Sci.*, **111**, E5272–E5281.
8. Klei, L., Sanders, S.J., Murtha, M.T., Hus, V., Lowe, J.K., Willsey, A.J., Moreno-de-luca, D., Yu, T.W., Fombonne, E., Geschwind, D. *et al.* (2012) Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism*, **3**(1), 1–13.
9. Keller, M.F., Saad, M., Bras, J., Bettella, F., Nicolaou, N., Sharma, M., Gibbs, J.R., Simo, J., Stefa, H., Heutink, P. *et al.* (2012) Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson' s disease. *Hum.Mol. Genet*, **21**, 4996–5009.
10. Loh, P.-R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., Schizophrenia Working Group of Psychiatric Genomics Consortium, de Candia, T.R., Lee, S.H., Wray, N.R. *et al.* (2015) Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. *Nat. Genet*, **47**, 1385–1392.
11. Lee, S.H., Yang, J., Goddard, M.E., Visscher, P.M. and Wray, N.R. (2012) Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics*, **28**, 2540–2542.
12. Gusev, A., Lee, S.H., Trynka, G., Finucane, H., Vilhja, B.J., Xu, H., Zang, C., Ripke, S., Bulik-sullivan, B., Stahl, E. *et al.* (2014) Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.*, **95**, 535–552 10.1016/j.ajhg.2014.10.004.
13. The Brian Consortium (2018) Analysis of shared heritability in common disorders of the brain. *Science*, **360**, 8757–8769.
14. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al.* (2015) Partitioning heritability by functional annota- tion using genome-wide association summary statistics. *Nat. Genet.*, **47**, 1228–1235.
15. Weissbrod, O., Flint, J. and Rosset, S. (2018) Estimating heritability and GeneticCorrelation in case control studies directly and with summary statistics. *Am. J. Hum. Genet.*, **103**, 89–99 10.1016/j.ajhg.2018.06.002.
16. Speed, D. and Balding, D.J. (2019) SumHer better estimates the SNP heritability of complex traits from summary statistics complex traits from summary statistics. *Nat. Genet.*, **51**, 277–284.
17. Visscher, P.M., Macgregor, S., Benyamin, B., Zhu, G., Gordon, S., Medland, S., Hill, W.G., Hottenga, J., Willemsen, G., Boomsma, D.I. *et al.* (2007) Genome partitioning of genetic variation for height from 11, 214 sibling pairs. *Am. J. Hum. Genet.*, **81**, 1104–1110.
18. Band, G., Rockett, K.A., Spencer, C.C.A., Kwiatkowski, D.P., Band, G., Si Le, Q., Clarke, G.M., Kivinen, K., Leffler, E.M., Rockett, K.A. *et al.* (2015) A novel locus of resistance to severe malaria in a region of ancient balancing selection. *Nature*, **526**, 253–257.
19. Timmann, C., Thye, T., Vens, M., Evans, J., May, J., Ehmen, C., Sievertsen, J., Muntau, B., Ruge, G., Loag, W. *et al.* (2012) Genome-wide association study indicates two novel resistance loci for severe malaria. *Nature*, **489**, 443–446.
20. Jallow, M., Teo, Y.Y., Small, K.S., Rockett, K.A., Deloukas, P., Clark, T.G., Kivinen, K., Bojang, K.A., Conway, D.J., Pinder, M. *et al.* (2009) Genome-wide and fine-resolution association analysis of malaria in West Africa. *Nat Genet*, **41**, 657–665 10.1038/ng.
21. Ravenhall, M., Campino, S., Sepu, N., Nadjm, B., Mtove, G., Wangai, H., Maxwell, C., Olomi, R., Reyburn, H., Drakeley, C.J. *et al.* (2018) Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-Eastern Tanzania. *PLoS Genet.*, **14**, e1007172 https://doi.org/10.1371/journal. pgen.1007172.
22. Malaria Genomic Epidemiology Network Consortium (2019) New insights into malaria susceptibility from the genomes of 17,000 individuals from Africa, Asia, and Oceania. *bioRxiv*. http://dx.doi.org/10.1101/535898.
23. Mayhew, A.J. and Meyre, D. (2017) Assessing the heritability of complex traits in humans: methodological challenges and opportunities. *Current Genom*, **18**, 332–340 10.2174/1389202918666170307161450.
24. Visscher, P.M., Brown, M.A., Mccarthy, M.I. and Yang, J. (2012) Five years of GWAS discovery. *Am. J. Hum. Genet.*, **90**, 7–24.
25. Flori, L., Sawadogo, S., Esnault, C., Fre, N., Fumoux, F. and Rihet, P. (2003) Linkage of mild malaria to the major histocompatibility complex in families living in Burkina Faso. *Hum. Mol. Genet.*, **12**, 375–378.
26. Brisebarre, A., Kumulungui, B., Sawadogo, S., Atkinson, A., Garnier, S., Fumoux, F. and Rihet, P. (2014) A genome scan for plasmodium falciparum malaria identifies quantitative trait loci on chromosomes 5q31, 6p21.3, 17p12, and 19p13. *Malar. J.*, **13**, 1–7.
27. Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. *et al.* (2015) The African genome variation project shapes medical genetics in Africa. *Nature*, **517**, 327–332.
28. Parker, M., Bull, S.J., Vries, J., Agbenyega, T., Doumbo, O.K. and Dominic, P. (2009) Ethical data release in genome-wide association studies in developing countries. *PLoSMed*, **6**, 11e1000143 10.1371/journal.pmed.1000143.
29. Achidi, E.A., Agbenyega, T., Allen, S., Amodu, O., Bojang, K., Conway, D., Corran, P., Deloukas, P., Djimde, A., Dolo, A. *et al.* (2008) A global network for investigating the genomic epidemiology of malaria. *Nature*, **456**, 732–737.
30. Band, G., Le, Q.S., Jostins, L., Pirinen, M., Kivinen, K., Jallow, M., Sisay-Joof, F., Bojang, K., Pinder, M., Sirugo, G. *et al.* (2013) Imputation-based meta-analysis of severe malaria in three African populations. *PLoS Genet*, 10.1371/journal.pgen.1003509.
31. Speed, D., Hemani, G., Johnson, M.R. and Balding, D.J. (2012) Improved heritability estimation from genome-wide SNPs. *Am. J. Hum. Genet.*, **91**, 1011–1021.
32. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de, P.I.W., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
33. Antonacci, F., Kidd, J.M., Marques-bonet, T., Ventura, M., Siswara, P., Jiang, Z. and Eichler, E.E. (2009) Characterization of six human disease-associated inversion polymorphisms. *Hum. Mol. Genet.*, **18**, 2555–2566 10.1093/hmg/ddp187.
34. Lee, S.H., Wray, N.R., Goddard, M.E. and Visscher, P.M. (2011) Estimating missing heritability for disease from genome-wide association studies. *Am. J. Hum. Genet.*, **88**, 294–305.

35. Jiang, X., Finucane, H.K., Schumacher, F.R., Schmit, S.L., Tyrer, J.P., Han, Y., Michailidou, K., Lesseur, C., Kuchenbaecker, K.B., Dennis, J. *et al.* (2019) Shared heritability and functional enrichment across six solid cancers. *Nat.Commun*, **10**, 431–454 10.1038/s41467-018-08054-4.
36. Gazal, S., Finucane, H.K., Furlotte, N.A., Loh, P. and Palamara, P.F. (2017) Linkage disequilibrium dependent architecture of human complex traits shows action of negative selection. *Nat Genet.*, **49**, 1421–1427.
37. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. and Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat Genet.*, **45**, 124–130 10.1038/ng.2504.
38. Sigova, A.A., Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-andre, V., Hoke, H.A. and Young, R.A. (2013) Resource super-enhancers in the control of cell identity and disease. *Cell.*, **155**, 934–947.
39. Hoffman, M.M., Ernst, J., Wilder, S.P., Kundaje, A., Harris, R.S., Libbrecht, M., Giardine, B., Ellenbogen, P.M., Bilmes, J.A., Birney, E. *et al.* (2013) Integrative annotation of chromatin elements from ENCODE data. *Nuc.Ac.Res.*, **41**, 827–841.
40. Lindblad-toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E. *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476–481.
41. Ward, L.D. and Kellis, M. (2012) Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Sci.*, **337**, 1675–1683.
42. Andersson, R., Gebhard, C., Miguel-escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461 10.1038/nature12787.
43. Pasaniuc, B., Zaitlen, N., Shi, H., Bhatia, G., Gusev, A., Pickrell, J., Hirschhorn, J., Strachan, D.P., Patterson, N. and Price, A.L. (2014) Fast and accurate imputation of summary statistics enhances evidence of functional enrichment. *Bioinformatics*, **30**, 2906–2914 10.1093/bioinformatics/btu416.
44. The 1000 Genomes Project Consortium (2011) A map of human genome variation from population scale sequencing. *Nature*, **467**, 1061–1073.

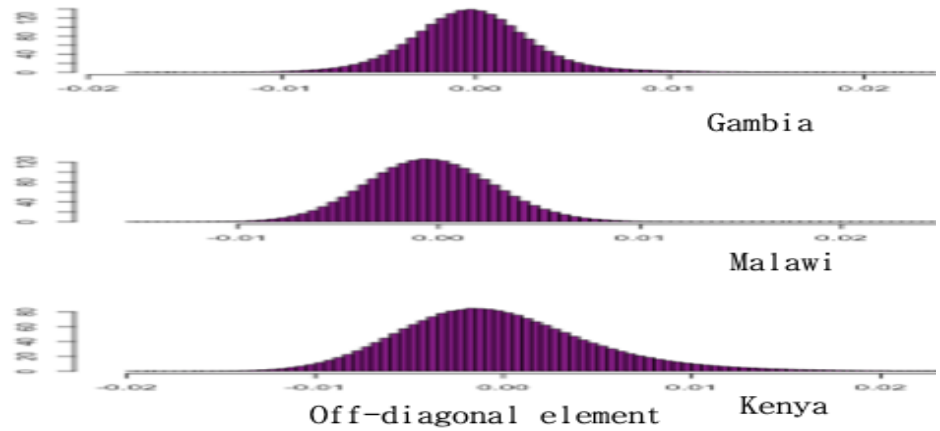
Supplementary Materials



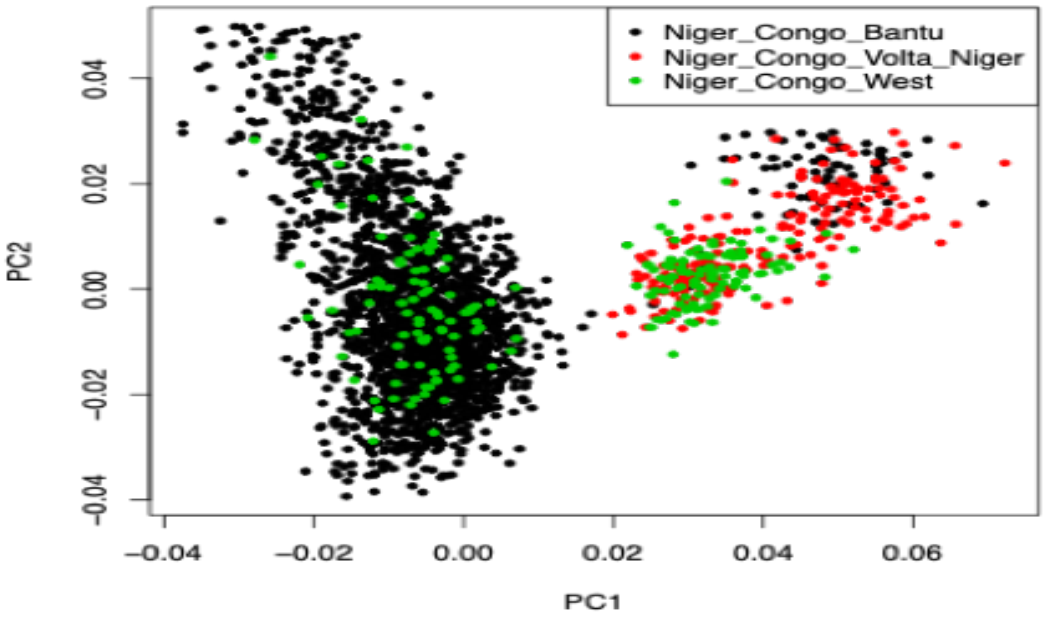
Supplementary Figure 1. SNP-heritability of severe malaria susceptibility and resistance at different QC-threshold from less stringent to more stringent (from left to right) determined GCTA. Error bars represent the 95% CI of the estimates. QC1: GWAS QC-filtered data, QC2:QC1+ samples with 5% relatedness removed, QC3: QC2+ SNPs with differential missing rates $1x10^{-3}$ removed, QC4:QC3+ SNPs with missingness proportion 2×10^{-2} removed, QC5: Q4+15 PCs as covariate, QC6: QC4+20 PCs, QC7=QC4+50PCs.



Supplementary Figure 2. SNP-heritability estimate per chromosome(y-axis) plotted against chromosome length(x-axis). The red line and blue line represent regressed SNP-heritability estimates obtained from joint GREML and separate analysis respectively. The 95% confidence interval around the slope of the regression model is represented by the grey shaded areas.



Supplementary Figure 3. Distribution of off-diagonal elements of the Genetic Relatedness Matrix (GRM) after removing closely related individuals (pairwise relatedness threshold >5%). As expected, the distribution was centred at zero suggesting that there was no close relatedness in all the datasets.



Supplementary Figure 4. We clustered our reference panel (quality filtered) in to geographic sub-regions in Africa using smart pca software. The left cluster represents east African population panel composed of 2112 samples and 22,473,268 SNPs. The right cluster represents the west African population panel composed of 380 samples and 18,919,068 SNPs.

Supplementary Table1. h^2_g of severe malaria resistance per chromosome estimated by GREML analysis implemented in GCTA

Chromosome	Chromosomal length (Mb)	h^2_g
1	249.3	0.019960
2	243.2	0.017271
3	198.0	0.014731
4	191.2	0.010840
5	180.9	0.03500
6	171.1	0.016039
7	159.1	0.000021
8	146.4	0.002449
9	141.2	0.003458
10	135.5	0.011938
11	135.0	0.028226
12	133.9	0.006632
13	115.2	0.01
14	107.3	0.01
15	102.5	0.001
16	90.4	0.006277
17	81.2	0.004484
19	78.1	0.001521
20	59.1	0.002942
21	63.0	0.016854
22	48.1	0.000001
Total		

4. CHAPTER FOUR: FUNCTIONAL ANALYSIS OF GENOME-WIDE DATASET FROM 17,000 INDIVIDUALS IDENTIFIES MULTIPLE CANDIDATE MALARIA RESISTANCE GENES ENRICHED IN MALARIA PATHOGENIC PATHWAYS

Summary

In this chapter, we applied various statistical functional analytic methods to the GWAS summary statistics (N=17,000) dataset of severe malaria resistance meta-analysed across eleven populations in malaria endemic regions in Africa, Oceania and Asia and identified plausible malaria resistance candidate genes and pathways. We performed population genetic structure of candidate malaria resistance gene in three malaria endemic populations (Gambia, Kenya and Malawi) and global populations consisting of 20 ethnic groups. We further applied gene burden and rare variant analysis on the genotype GWAS dataset of the three malaria endemic populations and identified rare variants that are associated with malaria resistance trait. Our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signalling elements and neuronal systems. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes that are enriched in overlapping pathways linked to severe malaria pathogenesis; highlighting, the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways.

Functional analysis of genome-wide dataset from 17000 individuals identifies multiple candidate malaria resistance genes enriched in malaria pathogenic pathways

Delesa Damena[1], Francis Agamah[1], Peter O. Kimathi[1], Ntumba E. Kabongo[1], Hundaol Girma[1], Wonderful T. Choga[1], Lemu Golassa[3] and Emile R. Chimusa[1,2*]

[1]Division of Human Genetics, Department of Pathology, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

[2]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Private Bag, Rondebosch, 7700, Cape Town, South Africa.

[3]Addis Ababa University, Aklilu Lema Institute of Pathobiology, P.O.Box 1176, Addis Ababa, Ethiopia

Abstract

Recent genome-wide association studies (GWASs) of severe malaria have identified several association variants. However, much about the underlying biological functions are yet to be discovered. Here, we systematically predicted plausible candidate genes and pathways from functional analysis of severe malaria resistance GWAS summary statistics(N=17,000) meta-analyzed across eleven populations in malaria endemic regions. We applied positional mapping, expression quantitative trait locus (eQTL), chromatin interaction mapping and gene-based association analyses to identify candidate severe malaria resistance genes. We performed network and pathway analyses to investigate their shared biological functions. We further applied rare variant analysis to raw GWAS datasets (N=11,000) of three malaria endemic populations including Kenya, Malawi and Gambia and performed various population genetic structures of the identified genes in the three populations and global populations.

Our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Our population genetic analysis revealed that the minor allele frequencies (MAF) of the single nucleotide polymorphisms (SNPs)

residing in the identified genes are generally higher in the three malaria endemic populations compared to global populations. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes; highlighting the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways.

Introduction

Malaria is still one of the global health problems with approximately 228 million cases and 405, 000 deaths in 2018 (1). African countries disproportionately carry the global burden of malaria accounting for 93% and 94% of cases and deaths, respectively (1). *P. falciparum* malaria is still one of the leading causes of child mortality in endemic regions, particularly in sub-Saharan Africa. According to the World Health Organization (WHO), malaria killed about 285,000 under five children in 2016 (2). About 10-20% of children who recover from severe malaria develop neurological sequalae and sub-optimal neuronal development (3). Severe malaria (SM) is defined as demonstration of asexual forms of the malaria parasites in the blood of a patient with a potentially fatal manifestation or complication of malaria in whom other diagnosis have been excluded (4). The SM complications include rapid progression to severe malarial anaemia (SMA), hypoglycaemia, cerebral malaria(CM), acidosis and death (4).

The global malaria eradication program, accelerated by the WHO led Roll Back Malaria (RBM) partnership, is largely focusing on scaling up the coverage of the available intervention strategies such as the distribution of long-lasting insecticide treated nets (LLINS), indoor residual insecticide spraying, intermittent treatment for pregnant women in high transmission settings and rapid diagnosis and effective treatments using artemisinin-based combination therapies (ACTs) (5). This led to the significant decline of malaria burden in many parts of the endemic regions. Despite the successes gained in reducing the global burden of malaria, the progress towards global malaria elimination is challenged by wide arrays of problems including decreased funding, lack of political commitments in some countries, emergence of drug resistant parasites and insecticide resistant mosquitoes and lack of effective vaccine among others (6).

P. falciparum has a complex life cycle alternate between vertebrate and female *Anopheles mosquito.* During its blood meal, the infected mosquitoes inoculates the transmissive form the parasite, the sporozoites, in to human skin. From the skin the sporozoites enter in to the blood

circulation or up-taken by the lymphatic system and invade liver (7). After maturation, the parasite buds off the hepatocytes and released in to the circulation in the form of merozomes containing hundreds of thousands of merozoites that infect erythrocytes (7). The erythrocyte stage also called blood stage lifecycle is complex multi-step process that involve repeated invasion, growth, replication and egress events (8). To create favourable environment for its survival and growth, the parasite remodels the infected erythrocyte (IE). The remodelling events include creation of a parasitophorous vacuolar membrane (PVM) that surrounds the parasite and modification of antigenic and structural properties of the IE by transporting a protein called *P. falciparum erythrocyte membrane protein 1 (PfEMP1)* to the surface (8). The clinical symptoms of the SM have been linked to the blood stage life cycle (8).

Even though SM is one of the commonest reasons for admission to hospital and is a major cause of hospital death in children aged 1-5 years in endemic areas, it constitutes only a small subset (1-2%) of the infected children as the majority of malaria infections is mild (9). It has been shown that such clinical variations is partly attributable to human genetic factors (10, 11). Thus, a comprehensive understanding of the human genetic causes of variation in malaria clinical outcome may potentially provide clues to design new intervention strategies such as therapeutics and vaccines which can facilitate the global malaria eradication program (12, 13).

Aiming at shedding more light to the genetic basis of severe *P. falciparum* malaria, several genome-wide association studies (GWASs) have been conducted in diverse malaria endemic populations over the last decade (14–18). The GWASs have replicated some of the well-known malaria susceptibility genomic risk loci including sickle cell (*HBB*) and *ABO* blood group loci and identified new variants in *ATP2B4* and Glycophorin regions (14–18). Due to the single-marker testing approach commonly used, single SNP-based GWASs may miss candidate variants with weak genetic effects and therefore, combining all effects from multiple variants within a gene and deconvoluting the interactions between genes underlying the malaria resistance trait may provide important insights in to underlying genetics (19). Today, a number of gene-based and pathway-level statistical analytic methods have recently been developed and successfully implemented in complex disease studies (20–23). These methods integrate functional information from advanced biological databases including the Genotype-Tissue Expression (GTEx)(24), Encyclopedia of DNA Elements (ENCODE) (25), Roadmap Epigenomics Project (26) and chromatin interaction

information (27). Owing to the fact that direct functional follow-up of several candidate causal variants and genes is expensive, application of such computational method to prioritize genes and their respective biological pathways are proven to be useful in complex diseases studies (20). Furthermore, gene and gene-set analysis can improve the study power by aggregating the joint effects of weakly associated markers, a common challenge in a polygenic trait studies (28).

Here, we implement several gene-set, pathway and network analytic methods on summary statistics of severe malaria GWAS from 17,000 individuals meta-analysed across eleven populations and systematically predicted plausible genes and pathways. We further performed rare variant analysis on raw GWAS dataset (N = ~11,000) of Kenya, Gambia and Malawi populations. Finally, we performed population genetic structure analysis of the identified genes in the three malaria endemic countries and across global populations. Established over the course of long co-evolution time, blood stage life cycle of the parasite constitutes the most extensive interplay between host and parasite genomes which leads to the clinical symptoms of SM. Therefore, our results suggest that severe malaria resistance is polygenic and attributed to multiple genes aggregated in pathogenic pathways linked to the erythrocyte stage lifecycle of *P. falciparum.*

Results

Functional mapping and annotations

We applied three functional mapping strategies to the severe malaria GWAS summary statistics (N=17,000 samples, ~17 million SNPs) meta-analysed across 11 malaria endemic populations in Africa, Asia and Oceania (see **Material and Methods**). We identified 19 lead SNPs out of 69 significant SNPs across 6 genomic loci (**Supplementary Data 1-3**). The genomic locus was defined as the region that contain independent lead SNPs and nominally significant SNPs ($p<0.05$) in linkage dis-equilibrium (LD) block with lead SNPs. An independent significant SNP was defined as a genome-wide significant SNP (P-value < 5e-8) within the genomic boundary of LD threshold of $r^2 > 0.6$. Lead SNPs were selected from independent significant SNPs at LD threshold of $r^2 > 0.1$. SNPs in close proximity (< 250 kb) were considered as a single locus and thus, each genomic locus can contain multiple independent significant SNPs and lead SNPs. These SNPs

were significantly enriched in ncRNA-intronic, intronic, intergenic and ncRNA-exonic regions (**Fig1.A**).

Our functional mapping strategies yielded a total of 57 protein-coding genes (**Table 1**, **Supplementary Data 4**). These include 29, 23 and 14 genes identified by eQTL mapping, chromatin interaction mapping and positional mapping, respectively (**Fig1.B**). Two genes including *ATP2B4* and *HBD* were identified by all the three gene mapping strategies while five genes including *GYPB, HBG2*, *TRIM6-TRIM34, OR51F2* and *TRIM68* were predicted by two of the three mapping strategies (**Supplementary Data 4**).

The identified genes were enriched in five cytogenic positions including 11p15 (p=2.65e-18), chr9q34 (p=4.63e-9) , chr4p31(p= 4.6e-8), chr1q32 (p=2.6e-7), chr3q26 (p=7.97e-4) (S1 Table). We noted that the majority (33%) of the identified genes were clustered on 11p15 (**Fig 2**). These include beta globin gene cluster: *HBB, HBD, HBG1, HBG2* and *HBE1;* Tripartite motif-containing (*TRIM)* family genes including *TRIM68, TRIM21*; and genes involved in olfactory receptors and G protein-coupled signalling (GPCR) such as: *CCKR OR51F2* and *OR51L* (**Supplementary Table 1**). About two third (13/19) of the genes in this locus are in eQTL and chromatin interactions (**Supplementary Data 4 and Supplementary Fig 1)**.

All the implicated genes in chr9q34 locus are located outside the genomic risk locus and were identified by eQTL mapping (**Supplementary Data 4, Supplementary Fig 2)**. These include surfeit gene cluster such as: *SURF2, SURF4, MED22* and *SURF6;* a metalloprotease gene, *ADAMTS13;* and a gene encoding *ORS* blood group system coding gene (*GBGT1)*. In the remaining enriched cytogenic positions, the known genes including *ATP2B4* (chr1q32); *FREM3, GYPE* and *GYPB* (chr4p31) were replicated. Other notable genes include *BTG2* a tumour suppressor gene on chr1q32 and *B3GALNT1* on chr3q26 (**Supplementary Data 4**).

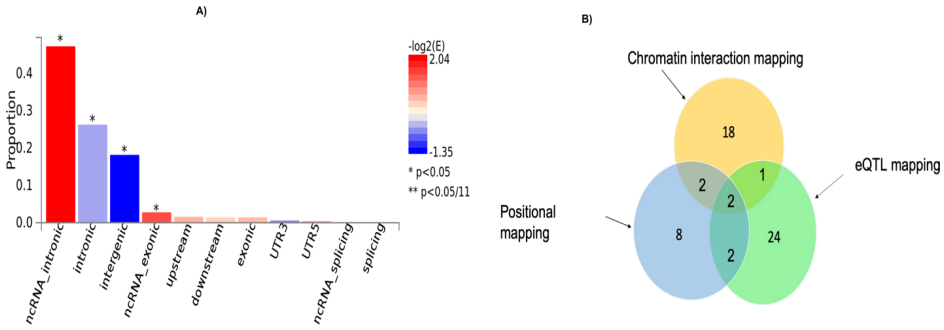


Fig.1. A) Proportion of malaria resistance GWAS SNPs in different genomic annotation categories. **B)** The number of genes identified by each of the three functional mapping strategies including positional mapping, eQTL and chromatin interactions. The intersection sections of the circles depict the number overlapped genes between the respective mapping strategies.

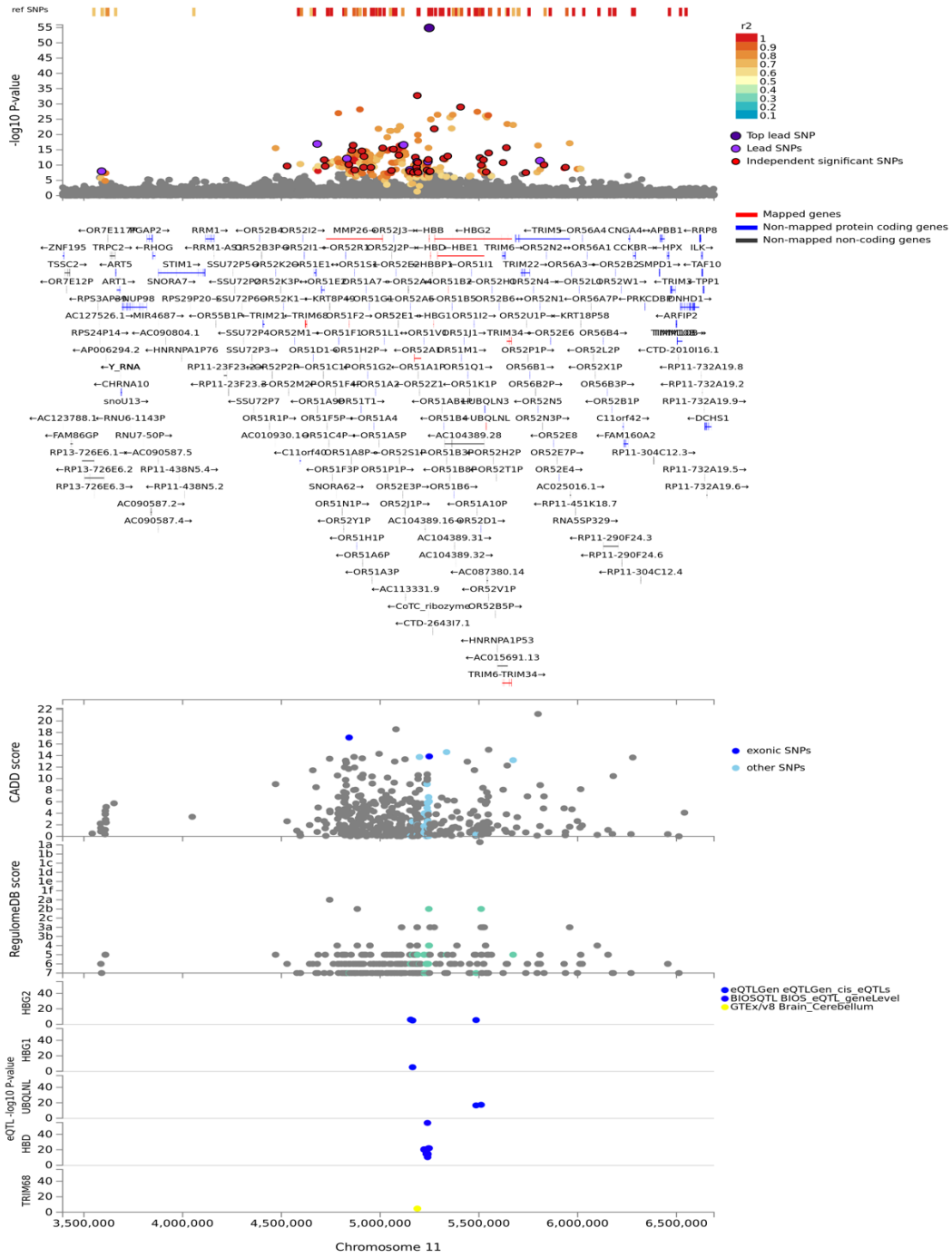


Fig. 2. Regional plot of severe malaria susceptibility GWAS locus on chromosome 11. Non-GWAS-tagged SNPs are shown at the top of the plot as rectangles since they do not have a P-value from the GWAS. Prioritized genes are highlighted in red. eQTLs are plotted per gene and coloured based on tissue types. CADD score, RegulomeDB score and eQTLs, SNPs which are not mapped to any gene are coloured grey.

Table 1. Fifty-seven severe malaria resistance candidate genes identified by eQTL mapping, chromatin interaction mapping and positional mapping strategies implemented in FUMA

Genes -Ensg	Symbols	Chr	Cytoband	Start	End	Biotype	Independent significant SNPs
ENSG00000160323	*ADAMTS13*	9	q34.2	136279478	136324508	protein coding	rs8176751; rs687621
ENSG00000197859	*ADAMTSL2*	9	q34.2	136397286	136440641	protein coding	rs8176751
ENSG00000179674	*ARL14*	3	q25.33	160394948	160396233	protein coding	rs116423146
ENSG00000058668	*ATP2B4*	1	q32.1	203595689	203713209	protein coding	rs4951370
ENSG00000169255	*B3GALNT1*	3	q26.1	160801671	160823172	protein coding	rs116423146
ENSG00000159388	*BTG2*	1	q32.1	203274619	203278730	protein coding	rs4951370
ENSG00000110148	*CCKBR*	11	p15.4	6280966	6293357	protein coding	rs113892119:rs28576676
ENSG00000133063	*CHIT1*	1	q32.1	203181955	203242769	protein coding	rs4951370
ENSG00000113758	*DBN1*	5	q35.3	176883609	176901402	protein coding	rs687621
ENSG00000122176	*FMOD*	1	q32.1	203309756	203320617	protein coding	rs4951370
ENSG00000183090	*FREM3*	4	q31.21	144498455	144621828	protein coding	rs201510180
ENSG00000109458	*GAB1*	4	q31.21	144257915	144395721	protein coding	rs111374053
ENSG00000148288	*GBGT1*	9	q34.2	136028340	136039332	protein coding	rs687621
ENSG00000250361	*GYPB*	4	q31.21	144917257	145061844	protein coding	rs201510180
ENSG00000197465	*GYPE*	4	q31.21	144792020	144826716	protein coding	rs34330779
ENSG00000244734	*HBB*	11	p15.4	5246694	5250625	protein coding	rs334
ENSG00000223609	*HBD*	11	p15.4	5253908	5256600	protein coding	rs334; rs4290259; rs79681613; rs113892119:rs28576676
ENSG00000213931	*HBE1*	11	p15.4	5289582	5526847	protein coding	rs145843585
ENSG00000213934	*HBG1*	11	p15.4	5269313	5271122	protein coding	rs7927066
ENSG00000196565	*HBG2*	11	p15.4	5274420	5667019	protein coding	rs145843585; rs183322782; rs148179286;rs7927066;rs11037724
ENSG00000203813	*HIST1H3H*	6	p22	27777842	27778314	protein coding	rs8176751
ENSG00000122188	*LAX1*	1	q32.1	203734304	203745361	protein coding	rs4951370
ENSG00000148297	*MED22*	9	q34.2	136205160	136214986	protein coding	rs8176751; rs687621
ENSG00000108960	*MMD*	17	q22	53469974	53499353	protein coding	rs8176751
ENSG00000167346	*MMP26*	11	p15.4	4726157	5013659	protein coding	rs141862673; rs145429724
ENSG00000169251	*NMD3*	3	q26.1	160822484	160971320	protein coding	rs116423146
ENSG00000184881	*OR51B2*	11	p15.4	5344541	5345582	protein coding	rs145843585
ENSG00000176925	*OR51F2*	11	p15.4	4842551	4843686	protein coding	rs141862673; rs145429724; rs113892119:rs28576676
ENSG00000176798	*OR51L1*	11	p15.4	5020213	5021160	protein coding	rs113892119:rs28576676
ENSG00000182070	*OR52A1*	11	p15.4	5172239	5207612	protein coding	rs116780407
ENSG00000228474	*OST4*	2	p23.3	27293340	27294641	protein coding	rs8176751
ENSG00000142657	*PGD*	1	p36.22	10458649	10480201	protein coding	rs687621
ENSG00000143850	*PLEKHA6*	1	q32.1	204187979	204346793	protein coding	rs4951370
ENSG00000163590	*PPM1L*	3	q25.33	160473390	160796695	protein coding	rs116423146
ENSG00000188783	*PRELP*	1	q32.1	203444956	203460480	protein coding	rs4951370
ENSG00000170955	*PRKCDBP*	11	p15.4	6340176	6341877	protein coding	rs28576676
ENSG00000160271	*RALGDS*	9	q34.2	135973107	136039301	protein coding	rs8176751; rs687621
ENSG00000148300	*REXO4*	9	q34.2	136271186	136283164	protein coding	rs8176751

ENSG00000080345	***RIF1***	**2**	**q23.3**	**152266397**	**152364527**	**protein coding**	**rs8176751**
ENSG00000170153	***RNF150***	**4**	**q31.21**	**141780961**	**142134031**	**protein coding**	**rs111374053**
ENSG00000136193	***SCRN1***	**7**	**p14.3**	**29959719**	**30029905**	**protein coding**	**rs687621**
ENSG00000160326	***SLC2A6***	**9**	**q34.2**	**136336217**	**136344259**	**protein coding**	**rs8176719; rs687621**
ENSG00000196542	***SPTSSB***	**3**	**q26.1**	**161062580**	**161090668**	**protein coding**	**rs116423146**
ENSG00000148290	***SURF1***	**9**	**q34.2**	**136218610**	**136223552**	**protein coding**	**rs8176751; rs687621**
ENSG00000148291	***SURF2***	**9**	**q34.2**	**136223428**	**136228045**	**protein coding**	**rs8176751**
ENSG00000148248	***SURF4***	**9**	**q34.2**	**136228325**	**136242970**	**protein coding**	**rs8176751**
ENSG00000148296	***SURF6***	**9**	**q34.2**	**136197552**	**136203235**	**protein coding**	**rs8176751; rs687621**
ENSG00000196628	***TCF4***	**18**	**q21.2**	**52889562**	**53332018**	**protein coding**	**rs687621**
ENSG00000132109	***TRIM21***	**11**	**p15.4**	**4406127**	**4414926**	**protein coding**	**rs28576676**
ENSG00000132274	***TRIM22***	**11**	**p15.4**	**5710919**	**5758319**	**protein coding**	**rs28576676**
ENSG00000258659	***TRIM34***	**11**	**p15.4**	**5640994**	**5665628**	**protein coding**	**rs183322782; rs148179286**
ENSG00000213186	***TRIM59***	**3**	**q25.33**	**160150233**	**160203561**	**protein coding**	**rs116423146**
ENSG00000121236	***TRIM6***	**11**	**p15.4**	**5617339**	**5634188**	**protein coding**	**rs28576676**
ENSG00000258588	***TRIM6-TRIM34***	**11**	**p15.4**	**5617955**	**5665628**	**protein coding**	**rs183322782; rs148179286; rs28576676**
ENSG00000167333	***TRIM68***	**11**	**p15.4**	**4619902**	**4629489**	**protein coding**	**rs10837488; rs4290259:rs113892119:rs28576676**
ENSG00000175518	***UBQLNL***	**11**	**p15.4**	**5535623**	**5537935**	**protein coding**	**rs11037724**
ENSG00000109445	***ZNF330***	**4**	**q31.21**	**142142041**	**142155851**	**protein coding**	**rs111374053**

Candidate genes identified by gene-based GWAS analysis

Taking the polygenic nature of severe malaria resistance trait in to consideration (11), we applied a pathway scoring algorithm(Pascal) (22) method that aggregates all SNPs within a gene and capture polygenic effects at the gene level (**see Material and Methods**). The Pascal analysis replicated 13 genes that were identified by our functional annotation methods in malaria genomic risk loci (**Supplementary Table 2**) and identified 125 additional genes across the genome (**Supplementary Data 5**). The genes with top scores outside genomic risk loci include *CSMD1* (p=1.58e-12) on chr8p23.2 and *RBFOX1* (p=9.76e-11) on chr16p13.3. *CSMD1* is an important regulator of complement activation and inflammation (29, 30) while *RBFOX1* encodes for an mRNA-splicing factor linked to autism spectrum disorders (31). A previous study in Tanzanian population reported association of variants in *RBFOX* gene with SM (17).

Other important genes identified by gene-based GWASs include neural adhesion molecules including *CNTN4* (p=3.88e-9) on chr3p26.3-p26.2; *PCSK5*(p=2.88e-11) on chr9q21.13; *CDH13* (p=4.19e-8) on chr16q23.3) and *TMEM132* (p=2.18e-8) on chr17q12 (**Supplementary Data 5**). These genes were reported to be linked to autism spectrum disorders and other neurodevelopmental conditions (32). Furthermore, protein kinases including *FLT4* (p=9.96e-8) on 5q35.3; and *PTPRT* (p=4.92e-7) on chr20q12-q13 and *PRKG1* ((p=1.2e-6) on 10q11.2-q21.1 were among the genes with top scores (**Supplementary Data 5**). *PTPRT* is a tyrosine phosphatase receptor involved in STAT3 pathway and was recently reported to be associated with mild malaria susceptibility in Benin populations (33). *PRKG1* is a cyclic guanosine monophosphate (GMP) dependent protein kinase which plays important roles in relaxation of vascular smooth muscle and inhibition of platelet aggregation (34). *FLT4* acts as a cell-surface receptor for vascular endothelial growth factor C (VEGFC) and vascular endothelial growth factor D (VEGFD), and plays an essential role in the development of the vascular network (35). It has been shown that VEGF and its receptor-related molecules are expressed in the brain tissues and reported to play protective during CM (36).

Gene-based rare variant association

Because rare variants are known to play role in the variation of most complex traits, we applied optimal unified sequence kernel association test (SKAT-O), which combines burden and variance-component analyses (37) to the raw genotype GWAS dataset of Gambia, Kenya and Malawi populations (**see Material and Methods**). The SKAT-O analysis identified a total of six and nine nominally significant genes in Gambia and Malawi populations, respectively. These include nine long intergenic non-protein coding RNAs (LincRNAs), *MIR4282*, *GLYR1, NDNF, EPB41L2, ATP8A1* and *WASF3* (**Supplementary Table 3**). However, none of these genes were significant after correction for multiple testing.

Functional networks and subnetworks of severe malaria resistance candidate genes

To investigate the functional interaction between all the candidate malaria resistance candidate genes identified in this study, we implemented network analysis (**see Materials and Methods**). Our global network generated 351 functional interactions between 268 genes. Topology analysis identified *ABO, HBB, HBD, HBE1, ATP2B4* as highly influential connector hub genes influencing at least two subnetworks/communities while *TRIM21* and *OR5F2* constituted independent communities. *MED22* and *OR551B6* constituted provincial hub genes (**Fig 3**).

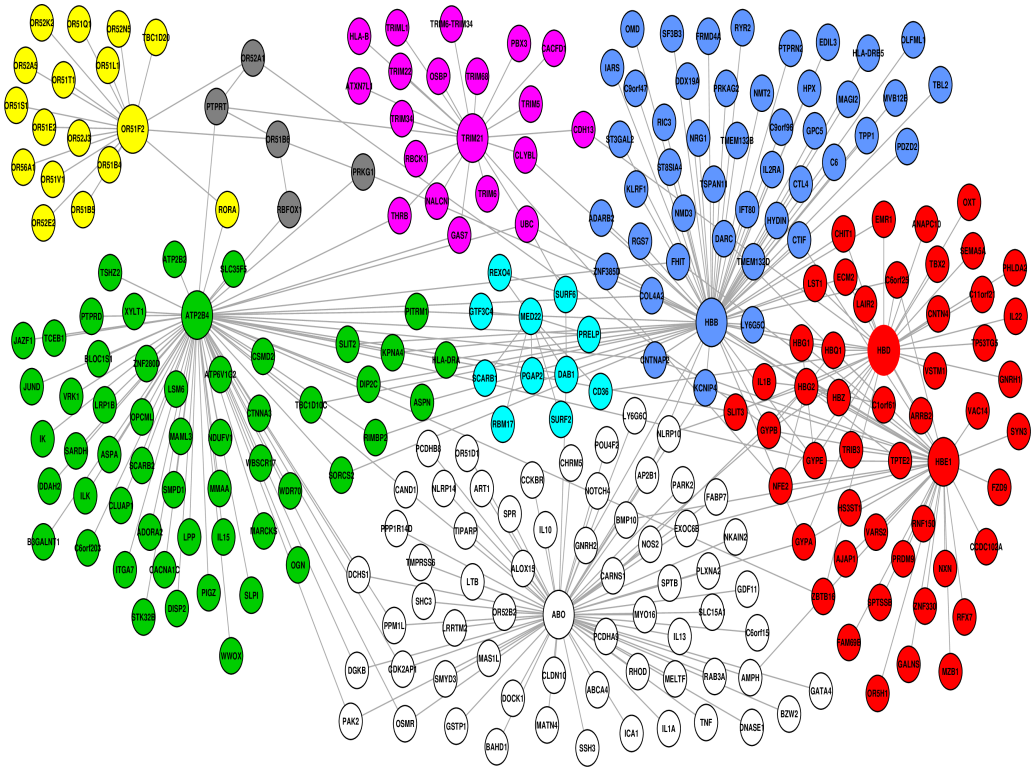


Fig. 3. Network generated from predominant severe malaria protective candidate genes, comprising of 351 interactions between 268 nodes. Topology analysis identified *ABO, HBB, HBD, HBE1, ATP2B4* as highly influential connector hub genes influencing at least two subnetworks/communities while *TRIM21* and *OR5F2* constituted independent communities. *MED22* and *OR551B6* constituted provincial hub genes.

Molecular functions and pathways of candidate severe malaria resistance genes

To test whether the genes predicted by the three functional mapping strategies overlapped in functional gene sets and pathways, we conducted gene enrichment analysis implemented in FUMA (20) using MsigDBc5 (38) gensets as background (**see Material and Methods**). The gene enrichment analysis identified several shared biological functions linked to erythrocyte-related pathways including three gene ontology (GO) cellular components, eight GO molecular functions and fourteen GO biological processes (**Table 2**). The implicated cellular components include haptoglobin-haemoglobin complex (p=7.6e-8), haemoglobin-complex(p=7.63e-8) and cytosolic-part (p= 6.73e-3). The enriched molecular functions include haptoglobin binding (p=4.87e-8), oxygen carrier activity(3.8e-7), oxygen binding (p=4.87e-5) and other activities related to haemoglobin functions. The shared biological activities include oxygen transport (p=4.22e-6), gas transport (p=4.87e-6), hydrogen peroxide catabolism (p=9.08e-5), protein hetero-oligomerization (p=2.86e-3), protein complex-oligomerization (p=4.23e-3), interferon gamma mediated signalling pathways (p=3.09e-2) and blood coagulation (p=3.09e-2).

In addition to genes within severe malaria genomic risk loci, we performed functional analysis and pathway analysis for the genes identified by the gene-based GWAS using Database for Annotation, Visualization and Integrated Discovery (DAVID) method (39) and Pascal (22), respectively(see Material and Methods). The DAVID analysis yielded eight functional categories, the majority of which are linked to malaria pathogenesis (**Table 3**) including GPCR signalling, membrane/transmembrane proteins, Na+/K+ transporting ATPases, cell adhesions, haemoglobin related functions, calcium signalling and actin binding activities.

Similarly, Pascal analysis implicated eleven significant pathways the majority of which are linked to malaria pathogenesis in RBCs, vasculatures and brain (**Fig 4A**). These include G protein-coupled receptor signalling (p=7.88e-15), haemostasis (p=4.52e-10), neuronal system (p=1.25 e-9), axon guidance (p=5.93e-8), calcium signalling (p=1.10e-7), chemical transmission across synapses (p=1.75 e-7), immune system (p=2.61e-6), signalling by Rho GTPase (p=1.45 e-5) and

Table 2. Gene enrichment results of functionally annotated genes in malaria genomic risk loci identified by FUMA method

GO terms	GeneSet	N. genes	N. enriched genes	P-value	adjusted P-value	Genes
Cellular components	Haptoglobin-haemoglobin complex	11	5	8.91e-11	7.63e-8	*HBB, HBD, HBG1, HBG2, HBE1*
	Haemoglobin complex	12	5	1.52e-10	7.63e-8	*HBB, HBD, HBG1, HBG2, HBE1*
	Cytosolic part	239	7	6.73e-6	2.25e-3	*HBB, HBD, HBG1, HBG2, HBE1, DBN1, SURF6*
Biological functions	Haptoglobin binding	10	5	4.87e-11	8.02e-8	*HBB, HBD, HBG1, HBG2, HBE1*
	Oxygen carrier activity	14	5	3.84e-10	3.16e-7	*HBB, HBD, HBG1, HBG2, HBE1*
	Oxygen binding	36	5	6.87e-8	3.77e-5	*HBB, HBD, HBG1, HBG2, HBE1*
	Molecular carrier activity	41	5	1.35e-7	5.56e-5	*HBB, HBD, HBG1, HBG2, HBE1*
	Oxidoreductase activity acting on peroxide as acceptor	56	5	6.66e-7	2.19e-4	*HBB, HBD, HBG1, HBG2, HBE1*
	Haemoglobin binding	7	3	8.69e-7	2.38e-4	*HBB, HBD, HBE1*
	Antioxidant activity	85	5	5.35e-6	1.26e-3	*HBB, HBD, HBG1, HBG2, HBE1*
	Tetrapyrrole binding	136	5	5.23e-5	1.08e-2	*HBB, HBD, HBG1, HBG2, HBE1*
Biological Processes	Oxygen transport	15	5	5.74e-10	4.22e-6	*HBB, HBD, HBG1, HBG2, HBE1*
	Gas transport	19	5	2.20e-9	8.10e-6	*HBB, HBD, HBG1, HBG2, HBE1*
	Hydrogen peroxide catabolic process	32	5	3.71e-8	9.08e-5	*HBB, HBD, HBG1, HBG2, HBE1*
	Antibiotic catabolic process	50	5	3.74e-7	6.88e-4	*HBB, HBD, HBG1, HBG2, HBE1*
	Drug catabolic process	108	6	8.05e-7	1.18e-3	*CHIT1, HBB, HBD, HBG1, HBG2, HBE1*
	Cofactor catabolic process	66	5	1.52e-6	1.87e-3	*HBB, HBD, HBG1, HBG2, HBE1*
	Protein hetero-oligomerization	133	6	2.73e-6	2.86e-3	*HBB, HBD, HBG1, HBG2, HBE1, HIST1H3H*
	Protein complex oligomerization	551	10	4.60e-6	4.23e-3	*TRIM21, HBB, HBD, HBG1, HBG2, HBE1, TRIM6, TRIM34, TRIM22, HIST1H3H*
	Antibiotic metabolic process	91	5	7.49e-6	6.11e-3	*HBB, HBD, HBG1, HBG2, HBE1*
	Cellular detoxification	107	5	1.65e-5	1.21e-2	*HBB, HBD, HBG1, HBG2, HBE1*
	Protein trimerization	54	4	2.00e-5	1.34e-2	*TRIM21, TRIM6, TRIM34, TRIM22*
	Detoxification	122	5	3.11e-5	1.91e-2	*HBB, HBD, HBG1, HBG2, HBE1*
	Interferon gamma mediated signalling pathway	70	4	5.60e-5	3.09e-2	*TRIM21, TRIM68, TRIM34, TRIM22*
	Coagulation	335	7	5.89e-5	3.09e-2	*HBB, HBD, HBG1, HBG2, HBE1, HIST1H3H, ADAMTS13*
	Oxygen transport	15	5	5.74e-10	4.22e-6	*HBB, HBD, HBG1, HBG2, HBE1*

Table 3: Functional categories of genes identified by gene-based GWAS analysis grouped using DAVID method

Functional group	Genes	Enrichment Score
GPCR signalling pathways and olfactory receptions	***OR51B6, FZD10, TMEM132C, OR51V1, OR51B2, OR52K2, SORCS2, OR56A1, OR52B4, TMEM132D, OR51E2, OR51T1, OR51B4, OR51B5, VSTM1, OR51F2, THSD7B***	**3.15**
Transmembrane protein and, Na+/K+ transporting ATPase	***TSPAN11, VSTM1, TMEM132D, SURF4, NKAIN2, TMEM132C, EVC, SLC35F3***	**3.07**
Tyrosine phosphatase, tyrosine kinase, cell adhesion molecule-like	***OPCML, PTPRD, PTPRT, CNTN5, FLT4, NTM, CNTN4, PTPRN2, VSTM1, PTPRS***	**2.51**
Haemoglobin related activities	***HBG2, HBE1, HBB, HBD***	**2.42**
Sodium leak and Potassium channel interacting protein	***KCNIP1, NALCN, KCNIP4, KCTD1***	**1.14**
zinc finger protein	***SMYD3, TSHZ2, ZNF385B, ZNF385D***	**0.61**
Actin binding LIM protein family and RAR-related orphan receptor A	***RORA, THRB, GLIS3, ABLIM2***	**0.45**
Calcium/calmodulin-dependent protein kinase, cGMP-dependent kinase and fms-related tyrosine kinase	***PRKG1, CAMK1D, FLT4, VRK1***	**0.44**

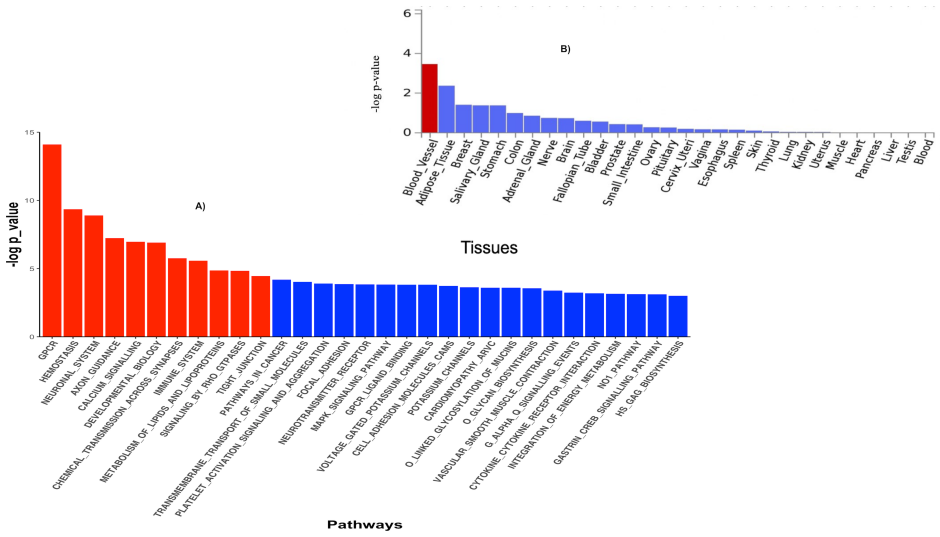


Fig. 4. A) Pathway scores obtained from pascal analysis. Significant Pathways at Bonferroni corrected P-value ≤ 0.05 are coloured in red. **B)** Tissue specific gene expression and shared biological functions of genes identified by pascal in GTEx v8 30 general tissue types. Input genes were tested against each of the precomputed differentially expressed sets using the hypergeometric test. Significant enrichment at Bonferroni corrected P-value ≤ 0.05 are coloured in red.

Population genetic structure of malaria resistance candidate genes

We noted that the minor allele frequencies (MAF) of the SNPs residing in the identified genes are generally higher in the three malaria populations compared to 20 ethnic groups (**Supplementary Fig 1 and 2**). We further observed that the proportion of pathogenic SNPs in a total of eighteen genes is much higher in the three malaria endemic populations compared to other populations (**Supplementary Data 6 and 7**). These include *TRIM* family genes such as *TRIM21, TRIM22, TRIM68, TRIM6-TRIM34 and TRIM34* in which the pathogenic SNP proportion ranges from 13.3-25%; olfactory receptors genes such as: *OR51B4, OR51B6, OR51B2, OR56A1, OR51L1, OR52K2 and OR51E2* in which the pathogenic SNP proportion ranges from (27.3 -100%).

The principal component analysis based on the SNPs residing in the identified genes effectively clustered all the populations according their ancestry (**Fig 5**).

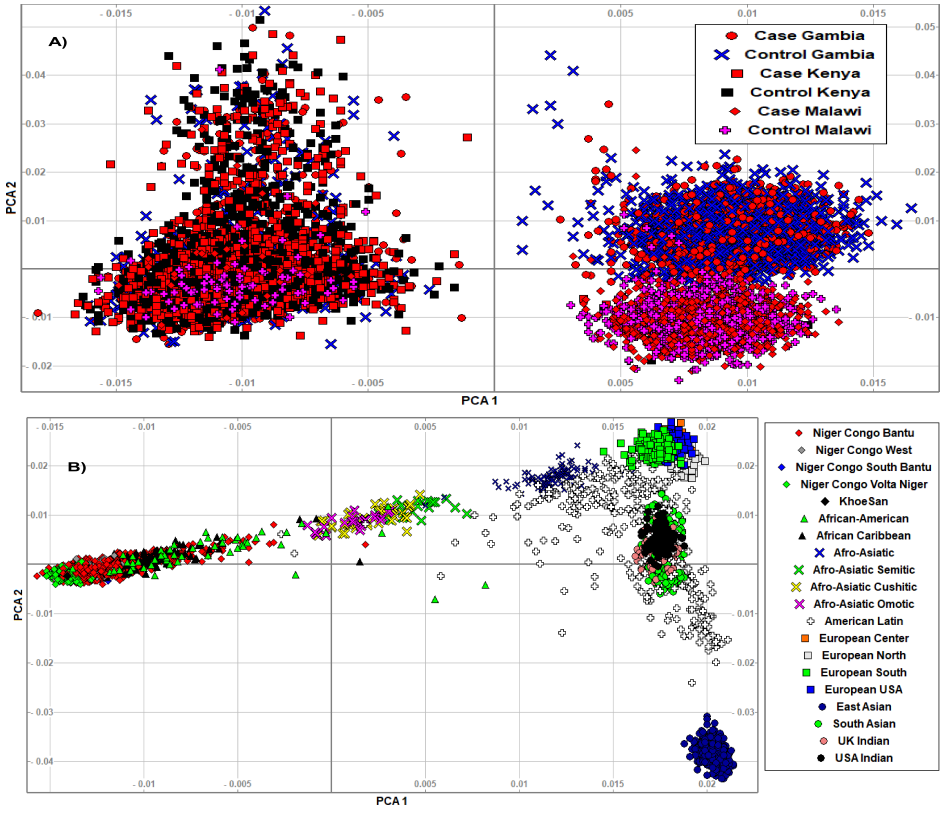


Fig .5. A) We clustered the merged malaria GWAS dataset containing only SNPs residing in the identified malaria resistance candidate genes (N= 10578 samples, 15,675 SNPs) using smartpca software. The three populations and their case/control status were indicated by different colours and symbols. **B).** We clustered the AGVP dataset containing only SNPs residing in the identified malaria resistance candidate genes (N=4932 samples, 93,5549 SNPs) in to sub-regions/populations using smartpca software. The populations were indicated by different colours and symbols.

Discussions

In this study, we applied statistical functional analytic method to the largest ever severe malaria susceptibility GWAS dataset and identified the well-known malaria resistance loci and a number of novel genes that can guide future functional experiments. We noted that severe malaria resistance is attributed to multiple genes and pathways linked to malaria pathogenesis during blood stage life cycle of the parasite including merozoite invasion, parasite growth, cytoadherence, and signal transduction. The genes that were identified by our three mapping strategies might have equal importance; genes that were identified by positional mapping may act at protein level through structural changes while the genes identified by eQTL and chromatin interactions exert their influences through quantitative changes at gene expression levels (20).

The fact that the functionally mapped genes are clustered on chromosome 11p15 is consistent with our recent work in which we reported the disproportionate concentration of SNP-heritability on chromosome eleven. This might reinforce the need for targeting this chromosome in the future severe malaria susceptibility studies (11). We noted that in addition to the sickle trait gene *(HBB)*, our mapping strategies identified other members of beta globin gene cluster that cause various forms of beta-thalassemia (*HBE1*, *HBD, HBG* and *HBG2*).

Our network analysis showed that all these genes constituted hub with which several other genes are connected; which reaffirm the importance of homeopathies in resistance against severe malaria. Beta-thalassemia and other hemoglobinopathies are thought to confer protection against severe malaria by suppressing the parasite growth and by mitigating associated pathogenic effects (40).

The proposed protective mechanisms of hemoglobinopathies against severe malaria has been extensively reviewed elsewhere (40, 41). In addition to the beta globin gene cluster, the Tripartite motif (*TRM)* containing gene family (*TRIM68* and *TRIM21*) identified in this locus are known to play critical role in down regulating Toll-like receptors(TLR)- and Rig-like receptors (RLR)-induced responses and protect from autoimmune and inflammations (42, 43). Mal-adapted inflammatory reactions is one of the hall mark pathogenic pathways in severe malaria (44, 45). The olfactory receptors super-family genes (*CCKR, OR51F2* and *OR51L*) identified in this locus might involve in G protein-coupled receptors(GPCR) signalling activities which is important in blood stage life cycle of *P. falciparum* (46). However, it is also possible that these genes were

detected because of their abundance and close proximity to with globin gene cluster(47). We also be noted that the genes in the *TRM* family and olfactory receptors super-family contain higher proportion of pathogenic SNPs in the three-malaria endemic population compared to the global populations. The majority of the well-known malaria protective genes have deleterious variants and were evolved under balancing selections [42].

Our eQTL mapping identified *ADAMTS13* gene on chr9q34 outside the malaria genomic risk locus which would not have been identified by the conventional SNP mapping approach (20). *ADAMTS13* is a zinc-containing metalloprotease enzyme that cleaves, von Willebrand factor vWF, a large protein derived from endothelial surface and megakaryocytes which plays a crucial role in basic haemostasis (48, 49). Following activation of endothelial cells, vWF is directly released in to plasma and basement membrane or is stored in Weibel–Palade bodies (WPBs) from where it is released by regulated secretion to promote adhesion of platelets at the sites of vascular injury and facilitate vascular healing (49). However, abnormal accumulations of vWF caused by deficiency of plasma *ADAMTS13* trigger intravascular platelet aggregation and micro thrombosis leading to a vascular disease, Thrombotic Thrombocytopenic Purpura (TTP) (48). Indeed, recent works have linked the platelet-mediated clumping of infected erythrocytes in microvasculature during cerebral malaria with increased level of VWF in plasma caused by mutations in *ADAMTS13* genes (50–52).

In the same genomic locus, our eQTL functional mapping identified surfeit gene cluster, metalloprotease genes linked to epithelial adhesions and blood coagulations. *SURF4* gene has been implicated in epithelial cell adhesion trait (53), *MED22* gene is linked to VWF factor/factor VIII level measurement (54) and *SURF6* has been implicated in epithelial ovarian cancer (55). Our network analysis showed that *MED22* forms a central hub to which the rest of surfeit gene cluster are connected. This may suggest that the greater importance of *MED22* gene compared with the other members of the cluster. However, further comprehensive studies are needed to better understand the association of these genes with SM resistance. In the remaining genomic risk loci, the well-known genes including *ATP2B4* (chr1q32), *FREM3, GYPE* and *GYPB* (chr4p31) were replicated and few additional genes were identified. The glycophorin gene cluster, *GYPA* and *GYPB* encode the MNS blood group system and are host-erythrocyte receptors for *P. falciparum*; suggesting that polymorphisms in these genes play protective role by interfering with the invasion

processes [55]. *ATP2B4* variants may impair the parasite lifecycle in erythrocyte by affecting intracellular Calcium homeostasis (18, 56).

Novel genes identified in the loci include *BTG2* on chr1q32 and *B3GALNT1* on chr3q26. *BTG2* is a tumour suppressor gene known to be linked to RBC related traits including MCHC level, RBC distribution and reticulocyte count (57). *B3GALNT1* encodes globoside blood group system which is determined by P antigen (58). Globoside/P antigen is the most abundant neutral glycolipid in the erythrocyte membrane and has been recognized as a cellular receptor for parvo-B19 virus (59). Individuals lacking this receptor are resistant to parvo-B19 virus and uro-pathogenic *E.coli* infections (59, 60). Further investigations are needed to establish the link between these genes and SM resistance.

We noted that the mapped genes share cellular components including haptoglobin binding, haemoglobin complex and cytosolic part and several overlapped molecular functions and biological processes linked with the blood stage life cycle of the parasite. *P. falciparum* spends most of its lifecycle within RBCs, where it undergoes multiple rounds of invasion, growth, replication and egress; causing the signs and symptoms of malaria (44, 61). The majority of the classical haemoglobin variants confer protection against severe malaria by restricting invasion process and intraerythrocytic growth of the parasite (40). Haptoglobin is an acute phase glycoprotein present in human plasma. It forms stable complexes with extracellular haemoglobin that is released from lysed RBCs and thereby curtail the haemoglobin-induced oxidative tissue damage (62).

P. falciparum ingest the host cell cytosol to obtain nutrients and space for growth in the RBCs (63). A recent study showed that the host cytosol uptake process is mediated by parasite's protein called *VPS45* (64). The fact that the identified candidate genes in this study were enriched in the cytosol part of the cellular component might suggest that these genes might arrest the nutrient uptake of the parasites and thereby confer protection against their pathogenic effects. In addition to haemoglobin related functions, some of the candidate genes were enriched in other pathways linked to malaria pathogenesis including blood coagulation related processes malaria (65, 66) and interferon gamma mediated signalling pathways (67, 68). This may suggest that the host genetic factors might interfere with parasite development and its pathogenic outcomes at multiple levels to confer protection the life-threatening form of malaria.

Our gene-based GWAS analysis replicated the well-known malaria resistance candidate genes in the genomic risk loci and identified several genes across the genome. Genes with top scores encode for different malaria relevant functions such as: regulation of inflammation (*CSMD1*), neural adhesion *(CNTN4, PCSK5, CDH13, TMEM132*), vascular epithelial development (*FLT4*) and protein kinases (*PTPRT, PRKG1*). Furthermore, functional analysis of the candidate genes yielded functional categories linked with the blood stage lifecycle of the parasite and associated pathologies. The top enriched functions include GPCR signalling, membrane/transmembrane proteins, [Na+/K+] transporting ATPases, Sodium leak and Potassium channel interacting proteins, cell adhesions molecules and Calcium/Calmodulin-dependent protein kinases (CDPKs). CDPKs have crucial functions in calcium signalling at various stages of the parasite's life cycle and is proposed to be one of the potential drug targets against malaria (69). It has been shown that *P. falciparum* infection activates host signalling pathway involving protein kinase C (PKC) (70). Similarly, host GPCR signalling pathways have been shown to play vital roles in invasion, intra-erythrocyte parasite development and egress processes (71, 72); suggesting the existence of substantial interactions between host membrane/transmembrane signalling and parasite signalling elements which might mediate the disease severity. Further studies are needed to decouple the host-parasite interface of signal transduction and explore the potential target for new therapeutics. Sodium leak and Potassium channel interacting proteins, [Na^+/K^+] transporting ATPases play critical role in maintaining electrochemical equilibrium in normal erythrocytes. However, upon invasion by trophoblast stage of the parasite, the ion pump-leak balance is perturbed; with increased leak rate and decreased pump rate resulting in a remarkable increase in [Na^+] and decrease in [K^+] in the erythrocyte cytosol (73, 74). This results in formation of a new permeability pathway (NPP) in the erythrocyte membrane which allow the transport of nutrients and waste products necessary for the parasite. The composition and physiological role of NPP has been reviewed elsewhere (75, 76). Studies have shown that both parasites driven proteins encoded by *clag3.1* and *clag3.2* (77, 78) and host benzodiazepine receptor mediate the formation of NPP (79, 80). Our result may suggest the existence of multiple host genes that are involved in mediating this process. Given that TPP can be a potential target for new therapeutics, further studies are needed to investigate the role of host and parasites genetics in mediating the channel and its pathophysiology.

In addition to the well-known pathways such as: haemostasis and immune system that play role in malaria resistance (13), our analysis identified novel pathways including cell adhesion molecules, Rho GTPase activities, tight junction, neuronal system and axon guidance. One of the key virulence factors of *P. falciparum* is its capacity to modify iRBCs to adhere to the endothelium of the vasculature and, thereby, sequester in capillaries and postcapillary venules in vital organs leading to severe disease manifestations (61). Adhesion phenotype is mediated by expression of *P. falciparum* erythrocyte membrane protein 1 (*PfEMP1*) on the iRBCs (81, 82). The binding of iRBCs with endothelium involve various adhesion molecules including CD36, ICAM-1, E-selectin and chondroitin sulfate A (CSA) that are variably expressed in different organs (83–85). The neural adhesion molecules identified in the current study might involve in receptor activities and their polymorphisms might play protective roles against SM. Furthermore, adhesion events have been shown to activate Rho kinase signalling pathway which is strongly implicated in various vascular diseases (86). The genes that are enriched in these pathways might provide protection against severe malaria by weakening the cytoadherence interactions and associated pathologies. Other genes that are enriched in neuronal system, axon guidance and tight junction might be linked with intra cerebral pathogenesis of SM (87). Furthermore, the candidate malaria resistance genes identified by gene-based GWAS were differential expressed in blood vessels; suggesting that the majority of the identified genes are likely counteract *P. falciparum* induced endothelial disfunctions in microvasculature and capillaries (44).

In conclusion, our functional mapping analysis identified 57 genes located in the known malaria genomic loci while our gene-based GWAS analysis identified additional 125 genes across the genome which can potentially guide future experimental studies. The identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes that are enriched in overlapping pathways linked to severe malaria pathogenesis; highlighting, the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways. Further experimental studies are needed to validate the findings in the current study.

Materials and Methods

Description of the study datasets

We accessed a previous severe malaria GWAS datasets (16) (N = ~11,000) of three African populations including Kenya, Gambia and Malawi from European Phenome Genome Archive (EGA) following the standard data access protocols outlined in (88, 89). Children with severe malaria cases were recruited on admission to hospital using definitions as per WHO guidelines for cerebral malaria (Blantyre coma score < 3 in children or Glasgow coma score < 11 in adults), severe malarial anaemia (haemoglobin < 5 g/100 mL or haematocrit < 15%) and other malaria-related symptoms (90). Control samples were obtained from representative of the ethnic groups of the cases or in some study sites from the local population (89). The samples were genotyped on Illumina Omni 2.5Marray and QC filtered as described in (14). In addition to the genotype dataset, we obtained a set of severe malaria susceptibility GWAS summary statistics (N=17,000) meta-analysed across eleven population in Africa, Oceania and Asia from (18). The dataset contained information on GWASs of individual study populations and their meta-analysis. We additionally accessed a merged data set from 1000 Genomes Project and African Genome Variation Project (AGVP)(91) of which 20 world-wide ethnic groups (**Supplementary Table 4**) have been grouped following ethno-linguistic information (92).

Functional mapping and annotations

We used the meta-analyzed malaria GWAS summary statistics (N=17,000 samples, 17million SNPs) across eleven populations(18) for functional mapping and annotations. We implemented FUMA(20), a pipeline that determines genomic risk loci and prioritize potential causal genes by incorporating information from multiple sources including GTEx (93), ENCODE (25), Roadmap Epigenomics Project (26) and chromatin interaction information (27). Briefly, a pre-calculated LD structure based on 1000 Genome version 3 of the African population was used to determine the risk loci and independent significant SNPS from the GWAS summary statistics data. Independent significant SNP was defined as a genome-wide significant SNP (P-value < 5e-8) within LD threshold of r^2 >0.6. Other nominally significant SNPs (p<0.05) that are in LD ($r^2 \leq 0.6$) with

independent significant SNPs were designated as candidate SNPs and included for annotation analysis. From identified independent significant SNPs, lead SNPs were selected at LD threshold of r^2 >0.1. SNPs in close proximity (< 250 k) were considered as a single locus and thus, each genomic locus can contain multiple independent significant SNPs and lead SNPs.

We then implemented three gene mapping strategies including positional mapping, expression Quantitative Trait Locus (eQTL) mapping, and chromatin interaction implemented in FUMA(20). Positional mapping was performed by ANNOVAR tool (94) using Ensembl (build 85; http://www.ensembl.org/) dataset. A maximum distance of 10kb window size upstream and downstream was used to map SNPs to genes. SNPs filtering was carried out based on CADD score(95), RegulomeDB score (96) and 15-core chromatin state (97).

eQTL mapping was performed for genes within 1 Mb of the most significant variant using datasets that contain eQTL information related to severe malaria such as brain and blood. These include:PsychENCODE(98), GTExv8 (93), BRAINEAC (99), DICE (100), eQTLGen (101) (and Blood eQTLbrowser (102) and scRNA_eQTLs(103). Chromatin interaction mapping was performed using dataset including Hi-C data of 21 tissues/cell types obtained from GSE87112. (104), Hi-C loops from Giusti-Rodriguez et al. 2019 (105) which contains pre-processed enhancer-promoter and promoter-promoter interactions based on Hi-C data for adult and foetal human brain samples(105), Hi-C based data from PsychENCODE (98) which is composed of Enhancer-Promoter links based on Hi-C and Promoter anchored Hi-C loops and Enhancer-Promoter correlations from FANTOM5 (106). We restricted our analysis to tissues related to severe malaria pathogenesis. Genes prioritized by any of these three strategies were considered as potential causal genes. To gain insights in to the biological functions of prioritized genes, we performed gene enrichment analysis using an hypergeometric test in which gene-sets obtained from MsigDB (38) and WikiPathways (107) were used as background genes. We further tested differential gene expression values on 54 tissues obtained from the GTEx (93) as indicted FUMA (20).

Gene-based genome-wide association analysis

Considering the polygenic nature of severe malaria susceptibility trait (11), we applied Pascal (22) , a gene-based GWAS analysis that can aggregate all SNPs within a gene and thus, capture modest effects. Briefly, Pascal sum of chi-squared statistics (SOCS) analysis was applied to nominally significant GWAS SNPs (p-value < 0.05) and to compute the corresponding gene scores (p-values) in 50kb upstream and downstream window. LD information for estimation of correlation structure was obtained from African dataset in 1000G phase 3 (108). We further categorized the prioritized genes in to different functional groups using DAVID tools (39). Significant genes were subjected to differential gene expression analysis implemented in FUMA software using 54 tissues obtained from the GTEx (93). Pathway scores were computed by combining the scores of genes that belong to the same gene-set. The analysis doesn't require a gene score threshold and incorporates weakly associated genes for pathway enrichment. Gene-fusion parameter was set to 1Mb so that all pathway-member genes within 1Mb apart were fused together. Genes in HLA-region and those containing more than 3000 SNPs were removed as outlined in Pascal documentations (22).

Gene burden and rare-variants association tests

Given that the GWAS assumption is based on common variant common disease hypothesis, GWAS approach always miss potential association signal from rare variants. To examine the contribution of rare variants, we applied optimal unified sequence kernel association test (SKAT-O), which combines burden and variance-component analyses (37) to the GWAS dataset of Gambia, Kenya and Malawi populations. Briefly, we aligned the VCF files including Gambia (N=4920 samples, 1.6 million SNPs), Malawi (N=2560 samples, 1.6 million SNPs) and Kenya (N=3143 samples, 1.6 million SNPs) to GWAS dataset to 1000 Genome v-3 reference haplotypes using Genotype Harmonizer (109) and removed SNPs with position and strand mismatches and phased using SHAPEITv2 (110). We performed imputation using impute 2 (111) and obtained ~ 20 million from each population. After removal SNPs with low genotype rate and imputation accuracy, we retained ~15,000 SNPs in each population. We then applied SKAT-O test to the quality filtered data following the procedure outlined in SKAT package (37).

Network analysis

We performed a network analysis to investigate the functional interaction between malaria resistance candidate genes. Briefly, we obtained functional interaction network from of the identified candidate malaria resistance genes using Multiple Association Network Integration Algorithm (geneMANIA) tool (112). Using this information, we computed network parameters including degree, betweenness and closeness centrality metrics to evaluate the topology of nodes(genes) and edges(interactions) in the network using networkX (113) and R igraph packages (114).

Closeness measures the average distance from the node to all other nodes in the network, indicating which nodes represent a greater "risk" (maximally close with lowest sum of edge weights) for eliciting other nodes. Betweenness measures the number of times that a node lies on the shortest path between two other nodes, indicating which nodes serve as a "hub" between other nodes (114). The degree of a node is described as the number of direct connections it has with other nodes within the network. Low degree nodes usually connect to nodes within their local community whereas high degree nodes usually extend to the neighbouring community (114). Using the centrality scores, we quantified node centrality to identify hub genes by investigating the contribution of the edges and the weight of the edges towards node centrality. The hubs genes make strong contributions to the subnetwork and/or global network integrity. Connector hubs and provincial hubs refers to nodes that link other nodes across different communities and local communities, respectively.

Population genetic structure of malaria-specific resistance candidate genes

We performed the population genetic structure of the identified genes in malaria endemic populations (Kenya, Gambia and Malawi) and global populations of 20 ethnic groups obtained from African Genome Variation Project (AGVP) (91). We mapped a total of 14,106,476 SNPs to the identified candidate genes using dbSNP database. We merged the quality filtered GWAS datasets of the three malaria endemic populations using PLINK software (115) and retained dataset that contains the SNPs that are mapped to the identified genes (N= 10578 samples, 15,675 SNPs). We performed basic quality control, removed structural variants and ambiguous SNPs

using PLINK software (115) as described elsewhere (116). We retained AGVP dataset containing only the SNPs that are mapped to the identified candidate genes (N=4932 samples, 93,5549 SNPs). We then partitioned these datasets in to a total 23 different Ethnic groups (20 from AGVP and 3 from malaria endemic populations) based on population or country labels information. We clustered the merged malaria GWAS dataset and the AGVP dataset containing only SNPs residing in the identified genes in to sub-regions/populations using smartpca software (117). For each ethnic group, we computed various population genetics analysis including minor allele frequency (MAF) at various bins and gene-specific in SNPs MAF by aggregating MAF of all associated SNPs within gene from dbSNPs. We finally computed proportion of pathogenic SNPs within the candidate genes using ANNOVAR software (94) by computing the number of SNPs reported to be pathogenic over total numbers associated SNPs within gene from dbSNPs.

Acknowledgements

This work was supported through the DELTAS Africa Initiative [grant 107740/Z/15/Z]. The DELTAS Africa Initiative is an independent funding scheme of the African Academy of Sciences (AAS)'s Alliance for Accelerating Excellence in Science in Africa (AESA) and supported by the New Partnership for Africa's Development Planning and Coordinating Agency (NEPAD Agency) with funding from the Wellcome Trust [grant 107740/Z/15/Z] and the UK government. The authors would also like to thank the National Research Foundation of South Africa for funding (NRF) [grant # RA171111285157/119056] and The Centre for High-Performance Computing (CHPC, www.chpc.ac.za) for computing resources This study makes use of data generated by MalariaGEN. A full list of the investigators who contributed to the generation of the data is available from www.malariagen.net. The views expressed in this publication are those of the author(s) and not necessarily those of AAS, NEPAD Agency, Wellcome Trust or the UK government.

Funding

DD, FA, PK, NK, HG and LG are funded by The Developing Excellence in Leadership and Genetics Training for Malaria Elimination in sub-Saharan Africa (DELGEME) program (grant #PD00217ML). EC is funded by NIH projects

Declaration of interest

The authors declare that they have no competing interests.

Authors contributions

DD designed, performed the data analysis and drafted the manuscript, FA, PK, NK, WT, HG and LG contributed in data analysis and revision of the manuscript, EC supervised the work

References

1. World Health Organization(WHO). World malaria report. 2019
2. World Health Organization(WHO) .World malaria report. 2018
3. Egan A, Keusch GT, Breman JG. The intolerable burden of malaria: a new look at the numbers.200. Am J Trop Med Hyg;64
4. World Health Organization(WHO). Severe Malaria. 2014;19: 7–131. doi:10.1111/tmi.12313
5. Dhiman S. Are malaria elimination efforts on right track ? An analysis of gains achieved and challenges ahead. IDP; 2019; 1–19.
6. Newby G, Bennett A, Larson E, Cotter C, Shretta R, Phillips AA, et al. The path to eradication : a progress report on the malaria-eliminating countries. The Lancet.2016; 1775–1784.
7. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature. 2002;415: 673–679. doi:10.1038/415673a
8. Cowman AF, Crabb BS. Invasion of Red Blood Cells by Malaria Parasites. Cell.2006; 755–766. doi:10.1016/j.cell.2006.02.006
9. Kevin M, Dayo F, Catherine W, Isiah M, Maria W, Victoria M, Et al. Indicators of life-threatening malaria in african children N Engl J Med.1995;332: 1399–1404.
10. Damena D, Denis A, Golassa L, Chimusa ER. Genome-wide association studies of severe P . falciparum malaria susceptibility : progress , pitfalls and prospects. BMC Med.Genom.; 2019; 1–14.
11. Damena D, Chimusa ER. Genome-wide heritability analysis of severe malaria resistance reveals evidence of polygenic inheritance.HMG. 2020;29: 168–176. doi:10.1093/hmg/ddz258
12. Teo Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association

analysis in Africa.Nat Rev Genet. 2013;11: 149–160.doi:10.1038/nrg2731.
13. Kwiatkowski DP. How Malaria Has Affected the Human Genome and What Human Genetics Can Teach Us about Malaria. Am J Hum Genet. 2005;77: 171–192. doi:10.1086/432519
14. Jallow M, Teo YY, Small KS, Rockett KA, Clark TG, Kivinen K, et al. Genmone-wide and fine-resolution association analysis of malria in West Africa. Nat Genet. 2010;41: 657–665. doi:10.1038/ng.388.Genome-wide
15. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012;489. doi:10.1038/nature11334
16. Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Si Le Q, Clarke GM, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature. 2015;526: 253–257. doi:10.1038/nature15390
17. Ravenhall M, Campino S, Sepu N, Nadjm B, Mtove G, Wangai H, et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. PloS Genet.2018; 1–15.
18. Gavin B, Quang S Le, Geraldine MC, Katja K, Christina H, Anna EJ, et al. Insights into malaria susceptibility using genome- wide data on 17,000 individuals from Africa, Asia and Oceania. Nat Commun. 2019;10: 5732. doi:doi.org/10.1038/s41467-019-13480-z |
19. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. Bioinformatics. 2015;32: 549–556. doi:10.1093/bioinformatics/btv619
20. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun; 2019: 1–10. doi:10.1038/s41467-017-01261-5
21. Leeuw CA De, Mooij JM, Heskes T, Posthuma D. MAGMA : Generalized Gene-Set Analysis of GWAS Data. PLoS Comput. Biol;2015; 1–19. doi:10.1371/journal.pcbi.1004219
22. Lamparter D, Marbach D, Rueedi R, Kutalik Z. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. PloS Comput Biol.2016; 1–20. doi:10.1371/journal.pcbi.1004714
23. Yoon S, Nguyen HCT, Yoo YJ, Kim J, Baik B, Kim S, et al. Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. Nucliec acids Res.2018;46. doi:10.1093/nar/gky175
24. The GTex Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science.2015;348: 648–661.
25. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature .2012;489:57–74 .
26. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. Nature.2015;48:317–330
27. Schmitt, A. D. et al. A compendium of chromatin contact maps reveals spatially active regions in the human genome. Cell Rep.2016; 17, 2042–2059
28. Dudbridge F. Polygenic Epidemiology. Genet Epidemiol. 2016;40: 268–272. doi:10.1002/gepi.21966
29. Sun PC, Uppaluri R, Schmidt AP, Pashia ME, Quant EC, Sunwoo JB, et al. Transcript Map of the 8p23 Putative Tumor Suppressor.Genomics; 2001;75: 17–25. doi:10.1006/geno.2001.6587
30. Lee AS, Rusch J, Lima AC, Usmani A, Huang N, Lepamets M, et al. Rare mutations in the complement regulatory gene CSMD1 are associated with male and female infertility. Nat

Commun. 2019; 10:1–16. doi:10.1038/s41467-019-12522-w
31. Hamada N, Ito H, Nishijo T, Iwamoto I, Morishita R. Essential role of the nuclear isoform of RBFOX1 , a candidate gene for autism spectrum disorders , in the brain development. Sci Rep. 2016; 1–19. doi:10.1038/srep30805
32. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, et al. Disruption of Contactin 4 (CNTN4) Results in Developmental Delay and Other Features of 3p Deletion Syndrome. Am. J. Hum. Genet. 2004;74: 1286–1293.
33. Milet J, Boland A, Luisi P, Sabbagh A, Sadissou I, Sonon P, et al. First genome-wide association study of non-severe malaria in two birth cohorts in Benin. Hum Genet.2019; 138:1341-1357
34. Mro´ zek K, Tanner M, Heinonen K, Bloomfield C. Molecular Cytogenetic Characterization of the KG-1 and KG-1a Acute Myeloid Leukemia Cell Lines by Use of Spectral Karyotyping and Fluorescence In Situ Hybridization. Gen, Chrom & Canc 2003;38:252: 249–252. doi:10.1002/gcc.10274
35. Aprelikova O, Pajusola K, Partanen J, Armstrong E, Alitalo R, Bailey SK, et al. FLT4) a Novel Class III Receptor Tyrosine Kinase in Chromosome. Canc Res.1992; 746–748.
36. Yeo TW, Lampah DA, Gitawati R, Tjitra E, Kenangalem E, Piera K, et al. Angiopoietin-2 is associated with decreased endothelial nitric oxide and poor clinical outcome in severe falciparum malaria.PNAS. 2008;105: 17097–17102.
37. Ionita-laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence Kernel Association Tests for the Combined Effect of Rare and Common Variants. Am J Hum Genet. The American Society of Human Genetics; 2013;92: 841–853. doi:10.1016/j.ajhg.2013.04.015
38. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3 . 0.Bionformatics. 2011;27: 1739–1740. doi:10.1093/bioinformatics/btr260
39. Jiao X, Sherman BT, Huang DW, Stephens R, Baseler MW, Lane HC, et al. DAVID-WS : a stateful web service to facilitate gene / protein list analysis. Bioinformatics.2012;28: 1805–1806. doi:10.1093/bioinformatics/bts251
40. Taylor SM, Cerami C, Fairhurst RM. Hemoglobinopathies: Slicing the Gordian Knot of Plasmodium falciparum Malaria Pathogenesis. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003327
41. Hedrick P. Population genetics of malaria resistance in humans. Heredity. 2011;10716: 283–304. doi:10.1038/hdy.2011.16
42. Wynne C, Lazzari E, Smith S, McCarthy EM, Nı´ Gabhann J, et al. TRIM68 Negatively Regulates IFN- b Production by Degrading TRK Fused Gene , a Novel Driver of IFN- b Downstream of Anti-Viral Detection Systems. PloS ONE 2014;9. doi:10.1371/journal.pone.0101503
43. Foss S, Bottermann M, Jonsson A, Sandlie I, James LC, Andersen JT. TRIM21 — From Intracellular Immunity to Therapy. Frontiers Immun.2019;10: 1–14. doi:10.3389/fimmu.2019.02049
44. Moxon A.Christopher, Matthew PG, McGuinness D, Danny AM, Matthias M. New Insights into Malaria Pathogenesis.Ann Rev. 2020;15:313-343
45. Milner DA. Malaria Pathogenesis. Cold Spring Harb Perspect Med. 2018; 1–11.
46. Mbengue A, Yam XY, Braun-breton C. Human erythrocyte remodelling during Plasmodium falciparum malaria parasite growth and egress. BJH.2012; 171–179. doi:10.1111/j.1365-2141.2012.09044.x

Plasmodium falciparum. Annu. Rev. Microbiol. 1997; 97–123.
64. Jonscher E, Flemming S, Schmitt M, Spielmann T, Jonscher E, Flemming S, et al. PfVPS45 Is Required for Host Cell Cytosol Uptake by Malaria Blood Stage Parasites.Cell Host Microb. 2019; 166–173. doi:10.1016/j.chom.2018.11.010
65. Francischetti IVOMB, Seydel KB, Monteiro RQ. Blood Coagulation , Inflammation , and Malaria. Microcirculation.2008; 81–107. doi:10.1080/10739680701451516
66. Sullivan JMO, Preston RJS, Regan NO, Donnell JSO. Emerging roles for hemostatic dysfunction in malaria pathogenesis.Blood. 2016;127: 2281–2289. doi:10.1182/blood-2015-11-636464.
67. Hunt NH, Ball HJ, Hansen AM, Khaw LT, Guo J, Bakmiwewa S, et al. Cerebral malaria : gamma-interferon redux. Frontiers Cell Infec.Microbiol. 2014;4: 1–12. doi:10.3389/fcimb.2014.00113
68. Tosevski V, Forman R, Muller W, Couper N. crossm Gamma Interferon Mediates Experimental Cerebral Malaria by Signaling within Both the Hematopoietic and Nonhematopoietic Compartments. Infect Immun. 2017;85: 1–13.
69. Ghartey-kwansah G, Yin Q, Li Z, Gumpper K, Sun Y, Yang R, et al. Calcium-dependent Protein Kinases in Malaria Parasite Development and Infection. Cell Transplant.2020; 1–12. doi:10.1177/0963689719884888
70. Sicard A, Semblat J, Doerig C, Hamelin R, Moniatte M, Dorin-semblat D, et al. Activation of a PAK-MEK signalling pathway in malaria parasite-infected erythrocytes. Cell Microbiol.2011;13: 836–845. doi:10.1111/j.1462-5822.2011.01582.x
71. Brochet M, Billker O. Calcium signalling in malaria parasites. Cell.Microbiol.2016; 12. doi:10.1111/mmi.13324
72. Millholland MG, Mishra S, Dupont CD, Love MS, Patel B, Shilling D, et al. Article A Host GPCR Signaling Network Required for the Cytolysis of Infected Cells Facilitates Release of Apicomplexan Parasites. Cell Host Microb.2013; 13:15–28.
73. Pillai AD, Addo R, Sharma P, Nguitragool W, Srinivasan P, Desai SA. Malaria parasites tolerate a broad range of ionic environments and do not require host cation remodelling. Mol Microbiol.2013;88: 20–34. doi:10.1111/mmi.12159
74. Desai SA. Why do malaria parasites increase host erythrocyte permeability ? Cell. 2014;30: 151–159.
75. Kiaran K. Ion Regulation in the Malaria Parasite. Biochem. J. 2014:457, 1–18
76. Kirk K, Lehane AM. Membrane transport in the malaria parasite and its host erythrocyte. Biochem. J. 2014;18: 1–18. doi:10.1042/BJ20131007
77. Nguitragool W, Bokhari AAB, Pillai AD, Rayavara K, Sharma P, Turpin B, et al. Malaria Parasite clag3 Genes Determine Channel-Mediated Nutrient Uptake by Infected Red Blood Cells. Cell.2011;145: 665–677. doi:10.1016/j.cell.2011.05.002
78. Ekland EH, Akabas MH, Fidock DA. Previews Taking Charge : Feeding Malaria via Anion Channels. Cell. 2011;145: 645–647. doi:10.1016/j.cell.2011.05.012
79. Bouyer G, Egée S, Thomas SL. Three types of spontaneously active anionic channels in malaria-infected. BCMD.2006;36: 248–254. doi:10.1016/j.bcmd.2006.01.005
80. Bouyer G, Cueff A, Gallagher PG, Thomas SLY. Erythrocyte peripheral type benzodiazepine receptor / voltage-dependent anion channels are upregulated by Plasmodium falciparum.Blood.2011;118: 2305–2312. doi:10.1182/blood-2011-01-329300.
81. Heatwoie VM, Wertheimer SP, Peterson DS, Ravetch JA, Weilems TE. The Large Diverse Gene Family w Encodes Proteins Involved in Cytoadherence and Antigenic Variation of

47. Alonso S, Lo S, Izagirre N. Overdominance in the Human Genome and Olfactory Receptor Activity. Mol. Biol. Evol. 2008;25:997–1001. doi:10.1093/molbev/msn049
48. Zheng X.Long. ADAMTS13 and von Willebrand Factor in Thrombotic Thrombocytopenic Purpura. Annu Rev Med. 2015; 66:211–225. doi:10.1146/annurev-med-061813-013241.
49. Sporn LA, Marder VJ, Wagner DD, Stel H V, Mourik JA Van. Differing Polarity of the Constitutive and Regulated Secretory Pathways for von Willebrand Factor in Endothelial Cells. J.Cell.Biol.1989;108: 1283–1289.
50. Adams Y, Kuhnrae P, Higgins MK, Ghumra A, Rowe JA. Rosetting Plasmodium falciparum-infected erythrocytes bind to human brain microvascular endothelial cells in vitro, demonstrating a dual adhesion phenotype mediated by distinct P. falciparum erythrocyte membrane protein 1 domains. Infect Immun. 2014;82: 949–959. doi:10.1128/IAI.01233-13
51. Hollestelle MJ, Mantey EA, Chakravorty SJ, Craig A, Akoto AO, Donnell JO, et al. von Willebrand factor propeptide in malaria : evidence of acute endothelial cell activation.BJH. 2006; 562–569. doi:10.1111/j.1365-2141.2006.06067.x
52. Kraisin S, Naka I, Patarapotikul J, Nantakomol D, Nuchnoi P. Association of ADAMTS13 polymorphism with cerebral malaria. Mal.J.2011; 1–8.
53. Ahsan M, Ek WE, Rask-andersen M, Karlsson T, Lind-thomsen A, Enroth S, et al. The relative contribution of DNA methylation and genetic variants on protein biomarkers for human diseases.PloS Genet. 2017; 13:1–24.
54. Sabater-Lleal M, Jennifer EH, Paul S de V, Jonathan M. Genome-Wide Association Transethnic Meta-Analyses Identifies Novel Associations Regulating Coagulation Factor VIII and von Willebrand Factor Plasma Levels. Circulations.2019; 139:620–635. doi:10.1161/CIRCULATIONAHA.118.034532
55. Chen K, Ma H, Li L, Zang R, Wang C, Song F, et al. Corrigendum : Genome-wide association study identifies new susceptibility loci for epithelial ovarian cancer in Han Chinese women. Nat Commun. 2014;5682: 6828. doi:10.1038/ncomms6828
56. Tiffert T, Lew VL, Ginsburg H, Krugliak M, Croisille L, Mohandas N. The hydration state of human red blood cells and their susceptibility to invasion by Plasmodium falciparum. Blood. 2005;105: 4853–4860. doi:10.1182/blood-2004-12-4948
57. Astle WJ, Elding H, Jiang T, Ouwehand WH, Butterworth AS, Soranzo N, et al. The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease Resource. Cell. 2017; 1415–1429. doi:10.1016/j.cell.2016.10.042
58. Hellberg Å, Poole J, Olsson ML. Molecular Basis of the Globoside-deficient P k Blood Group. J.Biol.Chem.2002;277: 29455–29459. doi:10.1074/jbc.M203047200
59. Kevin EB, Jonathan RH, Giorgio G, Stancie MA, Elton DL, Peggy M, et al. Resistance to Parvovirus B19 infection due to Lack of virus receptor(Erythrocyte Antigen). N Engl J Med. 1994;330: 1192–6.
60. Lund B, Lindberg FP, Baga M, Normark S. Globoside-Specific Adhesins of Uropathogenic Escherichia coli Are Encoded by Similar Trans-Complementable Gene Clusters.J.Bacteriol. 1985;162: 1293–1301.
61. Rowe JA, Claessens A, Corrigan RA, Arman M. Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Expert Rev Mol Med. 2009;11: e16. doi:10.1017/S1462399409001082
62. Kristiansen M, Graversen JH, Jacobsen C, Hoffman È, Law SKA, Sonne O, et al. Identintification of the haemoglobin scavenger receptor. Nature.2001;409: 1999–2002.
63. Francis SE, Sullivan DJ, Goldberg DE. hemoglobin metabolism in the malaria parasite

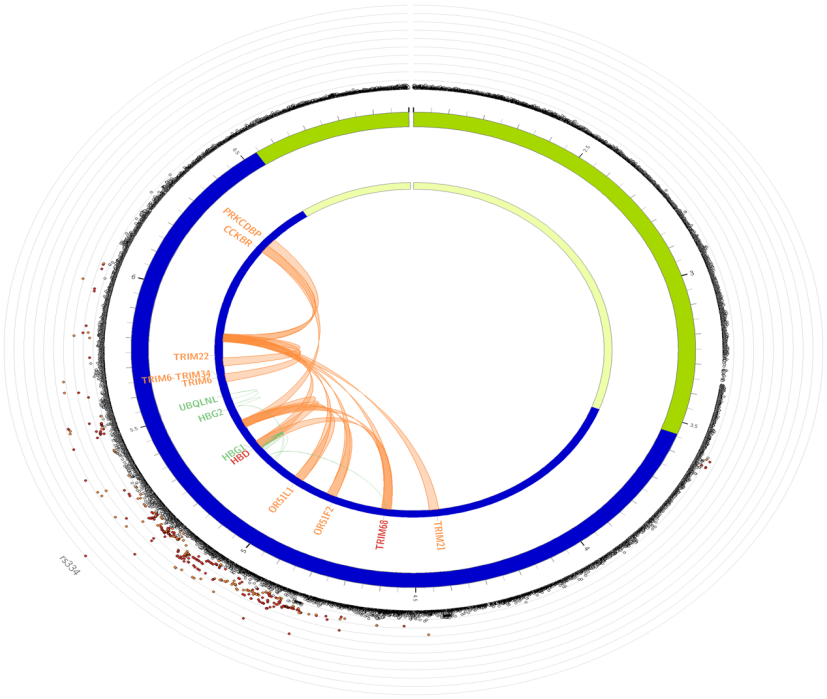
Plasmodium falciparum-Infected Erythrocytes.Cell. 1995;82: 89–100.
82. Smith JD, Craig AG, Roberts DJ, Hudson-taylor DE, Peterson DS, Pinches R, et al. Switches in Expression of Plasmodium falciparum VW Genes Correlate with Changes in Antigenic and Cytoadherent Phenotypes of Infected Erythrocytes. Cell. 1995;82: 101–110.
83. Raventos-suarez C, Kaul DK, Macalusot F, Nagel RL. Membrane knobs are required for the microcirculatory obstruction induced by Plasmodium falciparum-infected erythrocytes. Proc. Nadl. Acad. Sci. 1985;82: 3829–3833.
84. Rowe JA, Moulds JM, Newbold CI, Miller LH. mediated by a parasite-variant erythrocyte membrane protein and complement-receptor 1. Proc. Nadl. Acad. Sci. 1997;388: 292–295.
85. Fried M, Duffy PE. Adherence of Plasmodium falciparum to Chondroit i n Sulfate A in the Human Platen ta. Science.1996;272: 4–6.
86. Taoufiq Z, Gay F, Balvanyos J, Ciceron L, Tefit M, Lechat P, et al. Rho Kinase Inhibition in Severe Malaria : Thwarting Parasite-Induced Collateral Damage to Endothelia. JID.2008;197. doi:10.1086/528988
87. Nishanth G, Schlüter D. Blood – Brain Barrier in Cerebral Malaria : Pathogenesis and Therapeutic Intervention. Trends Parasitol. 2019;35: 516–528. doi:10.1016/j.pt.2019.04.010
88. Parker M, Bull SJ, Vries J De, Agbenyega T, Doumbo OK, Dominic P. Ethical Data Release in Genome-Wide Association Studies in Developing Countries. PloS Med.2009;6: 4–7. doi:10.1371/journal.pmed.1000143
89. Achidi EA, Agbenyega T, Allen S, Amodu O, Bojang K, Conway D, et al. A global network for investigating the genomic epidemiology of malaria. Nature. 2008;456: 732–737. doi:10.1038/nature07632
90. Trampuz A, Jereb M, Muzlovic I, Prabhu RM. Clinical review: Severe malaria. Crit Care. 2003;7: 315–323. doi:10.1186/cc2183
91. Gurdasani D, Carstensen T, Tekola-Ayele F, Pagani L, Tachmazidou I, Hatzikotoulas K, et al. The African Genome Variation Project shapes medical genetics in Africa. Nature. 2015;517: 327–332. doi:10.1038/nature139942
92. Michalopoulos S. The Origins of Ethnolinguistic Diversity Am. Econom Rev. 2012,. 2012;102: 1508–1539.
93. Guerini D, Pan B, Carafoli E. Expression, purification, and characterization of isoform 1 of the plasma membrane Ca2+ pump. Focus on calpain sensitivity. J Biol Chem. 2003;278: 38141–38148. doi:10.1074/jbc.M302400200
94. Wang K, Li M, Hakonarson H. ANNOVAR : Functional annotation of genetic variants from high-throughput sequencing data. 2010;38: 1–7. doi:10.1093/nar/gkq603
95. Kircher M, Witten DM, Jain P, Roak BJO, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nucl Ac Res,. 2014; doi:10.1038/ng.2892
96. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genom Res.2012; 1790–1797. doi:10.1101/gr.137323.112.
97. Ernst J, Kellis M. ChromHMM : automating chromatin- state discovery and characterization. Nat Method. 2012;10:1-2
98. Li M, Santpere G, Kawasawa YI,Evgrafov OV, Gulden FO, et al. Integrative functional genomic analysis of human brain development and neuropsychiatric risks. Science.2018; doi:10.1126/science.aat7615

99. Ramasamy A, Trabzuni D, Guelfi S, Varghese V, Smith C, Walker R, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat.Neuroscience. 2014; doi:10.1038/nn.3801
100. Schmiedel BJ, Singh D, Madrigal A,Valdovino-Gonzalez AL, White BM, Zapardiel-Gonzalo J, et al. Impact of Genetic Polymorphisms on Human Immune Resource Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. Cell.2018; 1701–1715.
101. Võsa U, Claringbould A, Westra H, Bonder MJ, Zeng B, Kirsten H, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta- analysis. bioRxiv 2018; 1–57. doi: https://doi.org/10.1101/447367 . this
102. Zhernakova D V, Deelen P, Vermaat M, Iterson M Van, Galen M Van, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. Blood.2016; doi:10.1038/ng.3737
103. Wijst MGP Van Der, Brugge H, Vries DH De, Deelen P, Swertz MA. Single-cell RNA sequencing identifies celltype- specific cis-eQTLs and co-expression QTLs.Nat.Genet. 2018. 50:493–497; https://doi.org/10.1038/s41588-018-0089-9
104. Schmitt AD, Hu M, Jung I, Lin Y, Barr CL. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. Cell.2016; 2042–2059.
105. Huckins LM, Dobbyn A, Ruderfer DM, Hoffman G, Wang W, Pardiñas AF, et al. Gene expression imputation across multiple brain Regions provides insights into schizophrenia risk. Nat.Genet.2019; 51: 659–674 https://doi.org/10.1038/s41588-019-0364-4
106. Andersson R, Gebhard C, Miguel-escalada I, Hoof I, Bornholdt J, Boyd M, et al. An atlas of active enhancers across human cell types and tissues. Nature. 2014.504:455–461 doi:10.1038/nature12787
107. Kutmon M, Riutta A, Nunes N, Hanspers K, Willighagen EL, Bohler A, et al. WikiPathways : capturing the full diversity of pathway knowledge. 2016; Nucl.Ac.Res. 44: 488–494. doi:10.1093/nar/gkv1024
108. The 1000 Genomes Project Consortium. Europe A map of human genome variation from population scale sequencing. Nature. 2011;467: 1061–1073. doi:10.1038/nature09534.A
109. Mostafavi S, Ray D, Warde-farley D, Grouios C, Morris Q. GeneMANIA : a real-time multiple association network integration algorithm for predicting gene function. Nucl.Ac.Res.2008;9: 1–15.
110. Hagberg AA, National LA, Alamos L. Exploring Network Structure , Dynamics , and Function using NetworkX. Proc. SciPy.2008; 11–15.
111. Csardi G, Tam'as N. The igraph software package for complex network research. IJCS
112. Deelen P, Bonder MJ, Velde KJ Van Der, Westra H, Winder E, Hendriksen D, et al. Genotype harmonizer : automatic strand alignment and format conversion for genotype data integration.BMC.Res.Not. 2014; 1–4.
113. Delaneau O, Marchini J, Zagury J. High-resolution whole-genome haplotyping using limited seed data.Nat.Method. 2013;9.
114. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009;5. doi:10.1371/journal.pgen.1000529
115. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J Hum Genet. 2007;81: 559–575. doi:10.1086/519795
116. Marees AT, Stringer S, Claire M, Derks EM. A tutorial on conducting genome - wide

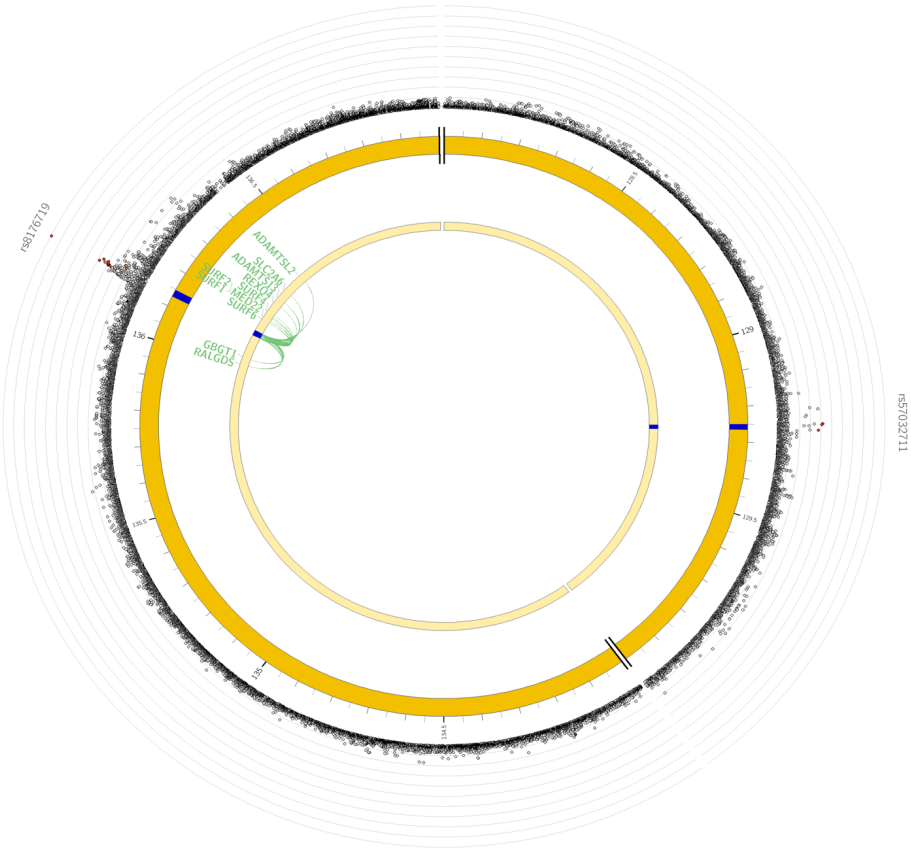
association studies : Quality control and statistical analysis. Int J Methods Psychiatr Res.2018; 1–10. doi:10.1002/mpr.1608

117. Patterson N, Price AL, Reich D. Population Structure and Eigenanalysis. PloS Genet.2006.2. doi:10.1371/journal.pgen.0020190

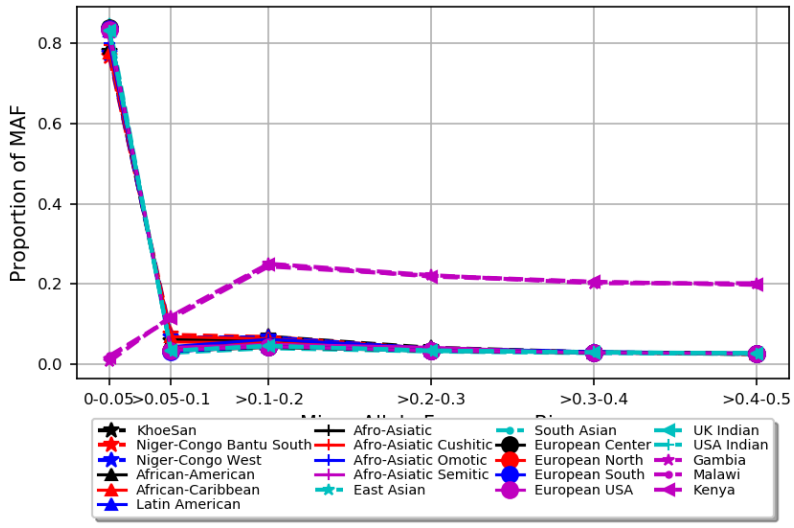
Supplementary Figures and Tables



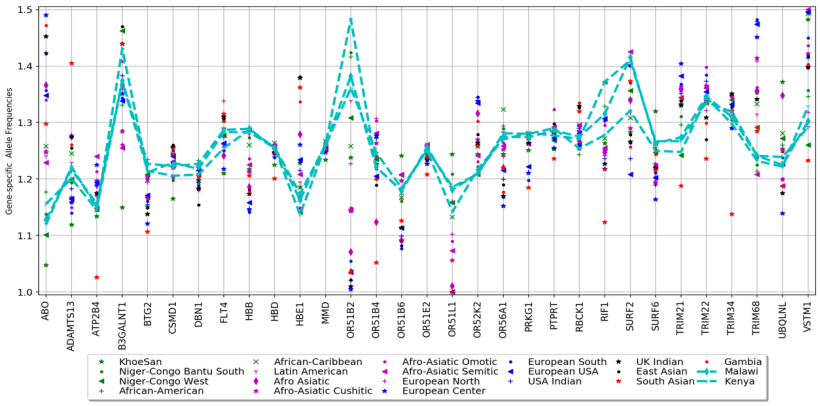
Supplementary Fig 1 . Chromatin interactions and eQTLs of severe malaria resistance candidate genes on chr 11 risk locus. The most outer layer is the Manhattan plot displaying SNPs with P-value < 0.05. Candidate SNPs are coloured based on the highest r^2 to one of the independent significant loci (red: $r^2 > 0.8$, orange: $r^2 > 0.6$). Other SNPs are coloured in grey. The outer circle is the chromosome coordinate and genomic risk loci are highlighted in blue. Genes mapped by either Hi-C or eQTLs are shown on the inner circle. Genes identified by chromatin interaction and eQTLs are coloured orange and green respectively while genes mapped by both are coloured red.



Supplementary Fig2. Chromatin interactions and eQTLs of severe malaria resistance candidate genes on chr 9 risk locus. The most outer layer is the Manhattan plot displaying SNPs with P-value < 0.05. Candidate SNPs are coloured based on the highest r^2 to one of the independent significant loci (red: $r^2 > 0.8$, orange: $r^2 > 0.6$). Other SNPs are coloured in grey. The outer circle is the chromosome coordinate and genomic risk loci are highlighted in blue. Genes mapped by either Hi-C or eQTLs are shown on the inner circle. Genes identified by chromatin interaction and eQTLs are coloured orange and green respectively while genes mapped by both are coloured red



Supplementary Fig 3. Minor allele frequency bins of SNPs mapped to severe malaria resistance candidate genes in three malaria endemic populations (Gambia, Malawi, Kenya) and global populations of 20 ethnic groups. Y-axis represent allele frequency, X-axis represent different MAF bins. Populations were represented by different colors and symbols



Supplementary Fig 4. Gene specific MAF of SNPs mapped to severe malaria resistance candidate genes in three malaria endemic populations (Gambia, Malawi, Kenya) and global populations of 20 ethnic groups. Y-axis represent gene specific allele frequency, X-axis represent genes. Populations were represented by different colors and symbols

Supplementary Table 1. Positional enrichment of genes identified by FUMA method using MsigDB genes as background

Ensg	entrezID	HUGO	symbol	chromosome	Cytoband	Start	end	Biotype
ENSG00000058668	493	*ATP2B4*	*ATP2B4*	1	q32.1	203595689	203713209	Protein coding
ENSG00000148296	6838	*SURF6*	*SURF6*	9	q34.2	136197552	136203235	Protein coding
ENSG00000148297	6837	*MED22*	*MED22*	9	q34.2	136205160	136214986	Protein coding
ENSG00000148290	6834	*SURF1*	*SURF1*	9	q34.2	136218610	136223552	Protein coding
ENSG00000148291	6835	*SURF2*	*SURF2*	9	q34.2	136223428	136228045	Protein coding
ENSG00000148248	6836	*SURF4*	*SURF4*	9	q34.2	136228325	136242970	Protein coding
ENSG00000132109	6737	*TRIM21*	*TRIM21*	11	p15.4	4406127	4414926	Protein coding
ENSG00000176925	119694	*OR51F2*	*OR51F2*	11	p15.4	4842551	4843686	Protein coding
ENSG00000244734	3043	*HBB*	*HBB*	11	p15.4	5246694	5250625	Protein coding
ENSG00000223609	3045	*HBD*	*HBD*	11	p15.4	5253908	5256600	Protein coding
ENSG00000196565	3048	*HBG2*	*HBG2*	11	p15.4	5274420	5667019	Protein coding
ENSG00000213931	3046	*HBE1*	*HBE1*	11	p15.4	5289582	5526847	Protein coding
ENSG00000184881	79345	*OR51B2*	*OR51B2*	11	p15.4	5344541	5345582	Protein coding

Supplementary Table 2. Candidate malaria resistance genes identified by both functional mapping and gene-based GWAS analysis

Position	N	N	P-value	adjusted P	Genes
chr11p15	297	19	8.86e-21	2.65e-18	*TRIM21, TRIM68, MMP26, OR51F2, OR51L1, OR52A1, HBB, HBD, HBG1, HBG2, HBE1, OR51B2, UBQLNL, TRIM6, TRIM6-TRIM34, TRIM34, TRIM22, CCKBR, PRKCDBP*
chr9q34	214	11	3.10e-11	4.63e-9	*RALGDS, GBGT1, SURF6, MED22, SURF1, SURF2, SURF4, REXO4, ADAMTS13, SLC2A6, ADAMTSL2*
chr4q31	67	6	4.61e-8	4.60e-6	*RNF150, ZNF330, GAB1, FREM3, GYPE, GYPB*
chr1q32	147	7	2.62e-7	1.96e-5	*CHIT1, BTG2, FMOD, PRELP, ATP2B4, LAX1, PLEKHA6*
chr3q26	61	3	7.97e-4	4.77e-2	*B3GALNT1, NMD3, SPTSSB*

Supplementary Table 3. Nominally significant genes obtained from burden and rare variant analysis using SKAT package

Population	Gene	chromosome	Cytoband	start	end	Biotype	P.value	N.Marker.Rare
Gambia	*LOC105377246*	4	q13.1	58,976,071	58,984,194	ncRNA	1.72e-4	29
	LOC100507464	8	q11.21	49,496,763	49,512,234	ncRNA	9.68e-05	36
	MIR4282	6	q13	72967687	72967753	lncRNA	4.83e-4	7
	GLYR1	16	P13.3	4803203	4847288	Protein coding	3.72e-4	139
	LINC00520	14	q22.3	55781132	55796731	lncRNA	9.70e-05	19
	LOC105371381	-	-	-	-	-	1.84e-4	15
Malawi	*ATP8A1*	4	P13	42408373	42657105	Protein coding	4.91e-4	244
	EPB41L2	6	q23.2	130839347	131063322	Protein coding	3.91e-4	207
	LOC105376161	9	q22.32	95,764,922	95,776,282	ncRNA	1.51e-4	9
	LOC105377731	5	q35.1	173,269,431	173,280,929	ncRNA	3.13e-4	29
	NDNF	4	q27	121035613	121073021	Protein coding	9.53e-05	28
	LINC00676	13	q13.34	109,728,274	109,730,034	lncRNA	3.40e-4	10
	LOC105370198	13	q14.2	48,108,123	48,166,377	ncRNA	4.87e-4	122
	LOC653786	-	-	-	-		4.53e-4	29
	WASF3	13	q12.13	26557683	26688948	Protein coding	2.54e-4	193

S2 Data 1. Severe malaria resistance genomic risk loci

GenomicLocus	uniqID	Lead SNPs	Chr	Position	P-value	Start	End
1	1:203652905:C:G	rs4951370	1	203652905	3.49E-08	203649423	203713887
2	3:160396863:C:T	rs116423146	3	160396863	1.64E-08	160362359	160397943
3	4:144587704:A:G	rs141274959	4	144587704	4.28E-10	143606326	145500856
4	9:129250119:A:G	rs57032711	9	129250119	1.31E-08	129250119	129264823
5	9:136132908:T:TC	rs8176719	9	136132908	1.16E-19	136128000	136148231
6	11:5248232:A:T	rs334	11	5248232	1.31E-55	3542398	6540324

S2 Data. Independent significant SNPs at r2>0.6 identified from severe malaria resistance GWASs

Genomic Locus	uniqID	rsID	Chr	Pos	P	nSNPs	nGWASSNPs
1	1:203652905:C:G	rs4951370	1	203652905	3.49E-08	76	64
2	3:160396863:C:T	rs116423146	3	160396863	1.64E-08	17	17
3	4:144501489:C:T	rs111374053	4	144501489	8.90E-10	141	8
3	4:144587704:A:G	rs141274959	4	144587704	4.28E-10	1	1
3	4:144653717:C:G	rs78251788	4	144653717	1.09E-08	16	2
3	4:144666678:A:C	rs149914432	4	144666678	7.79E-09	2	2
3	4:144747271:C:T	rs184497846	4	144747271	5.21E-10	53	34
3	4:144768545:A:G	rs2323342	4	144768545	1.36E-09	20	2
3	4:144778659:T:TA	rs201510180	4	144778659	2.29E-08	202	9
3	4:144783909:C:T	rs181143788	4	144783909	5.34E-09	171	24
3	4:144921625:G:T	rs111843057	4	144921625	1.68E-09	32	15
3	4:145064514:C:CTT	rs373218729	4	145064514	3.09E-09	1	1
3	4:145104286:C:T	rs149988318	4	145104286	1.48E-08	36	6
3	4:145124084:A:G	rs34330779	4	145124084	1.68E-08	1	1
4	9:129250119:A:G	rs57032711	9	129250119	1.31E-08	3	3
5	9:136131022:C:T	rs8176751	9	136131022	6.42E-10	19	17
5	9:136132908:T:TC	rs8176719	9	136132908	1.16E-19	1	1
5	9:136137065:A:G	rs687621	9	136137065	2.68E-09	36	6
6	11:3589834:C:T	rs76727905	11	3589834	1.02E-08	11	6
6	11:4528855:C:T	rs374510430	11	4528855	2.05E-10	7	3

6	11:4681473:G:T	rs188923516	11	4681473	1.19E-17	6	4
6	11:4716564:A:G	rs181023493	11	4716564	1.95E-12	9	6
6	11:4721821:C:T	rs12286023	11	4721821	2.39E-10	1	1
6	11:4831481:A:G	rs141862673	11	4831481	7.45E-13	72	70
6	11:4841199:A:G	rs145429724	11	4841199	6.33E-11	3	3
6	11:4856570:C:T	rs114980444	11	4856570	1.39E-15	7	7
6	11:4863120:A:G	rs146607281	11	4863120	3.11E-17	17	15
6	11:4871997:A:G	rs148547877	11	4871997	4.06E-09	10	9
6	11:4907203:C:T	rs75265800	11	4907203	2.39E-15	6	6
6	11:4916348:C:T	rs76444902	11	4916348	5.54E-10	9	7
6	11:4918226:A:G	rs184415744	11	4918226	1.44E-13	2	2
6	11:4953000:A:T	rs74052644	11	4953000	5.59E-10	4	4
6	11:5012800:A:G	rs142527103	11	5012800	2.71E-16	21	21
6	11:5057159:C:T	rs16908890	11	5057159	5.75E-09	28	28
6	11:5067053:C:G	rs139248000	11	5067053	6.35E-16	7	7
6	11:5094696:C:T	rs148559893	11	5094696	5.81E-14	20	18
6	11:5111447:A:G	rs147273018	11	5111447	2.86E-17	61	58
6	11:5118578:T:TA	rs201250014	11	5118578	2.45E-17	2	2
6	11:5150187:A:G	rs75366027	11	5150187	3.69E-09	8	8
6	11:5150751:A:G	rs7927066	11	5150751	1.08E-08	5	5
6	11:5168952:C:T	rs6578556	11	5168952	2.32E-08	3	3
6	11:5178410:G:T	rs189023578	11	5178410	2.75E-13	16	16
6	11:5184567:A:G	rs200657212	11	5184567	4.37E-10	1	1
6	11:5187079:A:G	rs7938610	11	5187079	8.80E-12	8	8
6	11:5187842:C:T	rs10837488	11	5187842	8.83E-09	3	3
6	11:5188027:G:T	rs183055323	11	5188027	1.79E-33	15	13
6	11:5190368:A:G	rs116780407	11	5190368	1.49E-11	28	28
6	11:5192015:C:G	rs73396897	11	5192015	2.62E-08	4	4
6	11:5236954:C:T	rs79681613	11	5236954	4.60E-09	9	9
6	11:5241282:C:G	rs4290259	11	5241282	6.79E-12	19	17
6	11:5243936:C:T	rs113850170	11	5243936	1.53E-12	2	2
6	11:5248232:A:T	rs334	11	5248232	1.31E-55	1	1
6	11:5249011:A:C	rs7119261	11	5249011	1.56E-08	1	1
6	11:5252794:A:G	rs12295158	11	5252794	1.02E-08	1	1
6	11:5273865:C:G	rs113892119	11	5273865	1.40E-22	6	3
6	11:5321510:C:CAGG	rs145843585	11	5321510	7.14E-13	10	10
6	11:5342001:A:G	rs11036798	11	5342001	1.13E-13	2	2
6	11:5407041:A:G	rs183322782	11	5407041	1.03E-29	29	14
6	11:5504744:AAA	rs11037724	11	5504744	3.71E-13	36	1
6	11:5511525:A:G	rs147525041	11	5511525	1.03E-10	17	16
6	11:5523004:C:T	rs150299658	11	5523004	1.87E-12	2	1
6	11:5536537:A:C	rs393044	11	5536537	1.74E-08	7	7

6	11:5549043:C:T	rs116554111	11	5549043	1.06E-14	3	3
6	11:5622634:A:G	rs141046578	11	5622634	1.55E-11	2	1
6	11:5639076:T:TA	rs141872087	11	5639076	2.06E-16	12	2
6	11:5737256:C:T	rs186977392	11	5737256	2.87E-08	1	1
6	11:5808878:A:C	rs148179286	11	5808878	3.74E-12	8	4
6	11:5829334:A:G	rs28576676	11	5829334	8.60E-11	1	1
6	11:5938210:A:G	rs183089109	11	5938210	5.33E-10	7	7

S3 Data. Lead SNPs identified from independent significant SNPs of severe malaria resistance GWAS

GenomicLocus	uniqID	rsID	chromosome	Position	P-value	nIndSigSNPs
1	1:203652905:C:G	rs4951370	1	203652905	3.49E-08	1
2	3:160396863:C:T	rs116423146	3	160396863	1.64E-08	1
3	4:144587704:A:G	rs141274959	4	144587704	4.28E-10	3
3	4:144768545:A:G	rs2323342	4	144768545	1.36E-09	6
3	4:144783909:C:T	rs181143788	4	144783909	5.34E-09	2
3	4:144921625:G:T	rs111843057	4	144921625	1.68E-09	2
3	4:145064514:C:CTT	rs373218729	4	145064514	3.09E-09	1
3	4:145104286:C:T	rs149988318	4	145104286	1.48E-08	3
4	9:129250119:A:G	rs57032711	9	129250119	1.31E-08	1
5	9:136132908:T:TC	rs8176719	9	136132908	1.16E-19	3
6	11:3589834:C:T	rs76727905	11	3589834	1.02E-08	1
6	11:4681473:G:T	rs188923516	11	4681473	1.19E-17	22
6	11:4831481:A:G	rs141862673	11	4831481	7.45E-13	8
6	11:5118578:T:TA	rs201250014	11	5118578	2.45E-17	9
6	11:5150187:A:G	rs75366027	11	5150187	3.69E-09	15
6	11:5187079:A:G	rs7938610	11	5187079	8.80E-12	8
6	11:5241282:C:G	rs4290259	11	5241282	6.79E-12	4
6	11:5248232:A:T	rs334	11	5248232	1.31E-55	13
6	11:5808878:A:C	rs148179286	11	5808878	3.74E-12	24

S4 Data. Prioritized genes from severe malaria resistance GWAS by three functional mapping strategies

Ensg	symbol	Chr	Start	end	eqtl Mapts	ciMap	IndSigSNPs	GenomicLocus
ENSG00000142657	PGD	1	10458649	10480201	No	No	rs687621	5
ENSG00000133063	CHIT1	1	203181955	203242769	NA	Yes	rs4951370	1
ENSG00000159388	BTG2	1	203274619	203278730	NA	Yes	rs4951370	1
ENSG00000122176	FMOD	1	203309756	203320617	NA	Yes	rs4951370	1
ENSG00000188783	PRELP	1	203444956	203460480	NA	Yes	rs4951370	1
ENSG00000058668	ATP2B4	1	203595689	203713209	Yes	Yes	rs4951370	1
ENSG00000122188	LAX1	1	203734304	203745361	Yes	No	rs4951370	1
ENSG00000143850	PLEKHA6	1	204187979	204346793	NA	Yes	rs4951370	1
ENSG00000228474	OST4	2	27293340	27294641	Yes	No	rs8176751	5
ENSG00000080345	RIF1	2	152266397	152364527	Yes	No	rs8176751	5
ENSG00000213186	TRIM59	3	160150233	160203561	NA	Yes	rs116423146	2
ENSG00000179674	ARL14	3	160394948	160396233	NA	No	rs116423146	2
ENSG00000163590	PPM1L	3	160473390	160796695	NA	Yes	rs116423146	2
ENSG00000169255	B3GALNT1	3	160801671	160823172	NA	Yes	rs116423146	2
ENSG00000169251	NMD3	3	160822484	160971320	NA	Yes	rs116423146	2
ENSG00000196542	SPTSSB	3	161062580	161090668	NA	Yes	rs116423146	2
ENSG00000170153	RNF150	4	141780961	142134031	NA	Yes	rs111374053	3
ENSG00000109445	ZNF330	4	142142041	142155851	NA	Yes	rs111374053	3
ENSG00000109458	GAB1	4	144257915	144395721	NA	No	rs111374053	3
ENSG00000183090	FREM3	4	144498455	144621828	Yes	No	rs201510180	3
ENSG00000197465	GYPE	4	144792020	144826716	Yes	No	rs34330779	3
ENSG00000250361	GYPB	4	144917257	145061844	Yes	No	rs201510180	3
ENSG00000113758	DBN1	5	176883609	176901402	Yes	No	rs687621	5
ENSG00000203813	HIST1H3H	6	27777842	27778314	Yes	No	rs8176751	5
ENSG00000136193	SCRN1	7	29959719	30029905	Yes	No	rs687621	5
ENSG00000160271	RALGDS	9	135973107	136039301	Yes	No	rs8176751;rs687621	5
ENSG00000148288	GBGT1	9	136028340	136039332	Yes	No	rs687621	5
ENSG00000148296	SURF6	9	136197552	136203235	Yes	No	rs8176751;rs687621	5
ENSG00000148297	MED22	9	136205160	136214986	Yes	No	rs8176751;rs687621	5
ENSG00000148290	SURF1	9	136218610	136223552	Yes	No	rs8176751;rs687621	5
ENSG00000148291	SURF2	9	136223428	136228045	Yes	No	rs8176751	5
ENSG00000148248	SURF4	9	136228325	136242970	Yes	No	rs8176751	5
ENSG00000148300	REXO4	9	136271186	136283164	Yes	No	rs8176751	5
ENSG00000160323	ADAMTS13	9	136279478	136324508	Yes	No	rs8176751;rs687621	5
ENSG00000160326	SLC2A6	9	136336217	136344259	Yes	No	rs8176719;rs687621	5
ENSG00000197859	ADAMTSL2	9	136397286	136440641	Yes	No	rs8176751	5
ENSG00000132109	TRIM21	11	4406127	4414926	NA	Yes	rs28576676	6
ENSG00000167333	TRIM68	11	4619902	4629489	Yes	Yes	rs10837488;rs4290259:rs113892119:rs28576676	6
ENSG00000167346	MMP26	11	4726157	5013659	NA	No	rs141862673;rs145429724	6

ENSG00000176925	OR51F2	11	4842551	4843686	NA	Yes	rs141862673;rs145429724;rs113892119:rs28576676	6
ENSG00000176798	OR51L1	11	5020213	5021160	NA	Yes	rs113892119:rs28576676	6
ENSG00000182070	OR52A1	11	5172239	5207612	NA	No	rs116780407	6
ENSG00000244734	HBB	11	5246694	5250625	NA	No	rs334	6
ENSG00000223609	HBD	11	5253908	5256600	Yes	Yes	rs334;rs42902 59;rs79681613 ;rs113892119: rs28576676	6
ENSG00000213934	HBG1	11	5269313	5271122	Yes	No	rs7927066	6
ENSG00000196565	HBG2	11	5274420	5667019	Yes	No	rs145843585;rs183322782;rs148179286;rs7927066;rs11037724	6
ENSG00000213931	HBE1	11	5289582	5526847	NA	No	rs145843585	6
ENSG00000184881	OR51B2	11	5344541	5345582	NA	No	rs145843585	6
ENSG00000175518	UBQLNL	11	5535623	5537935	Yes	No	rs11037724	6
ENSG00000121236	TRIM6	11	5617339	5634188	NA	Yes	rs28576676	6
ENSG00000258588	TRIM6-TRIM34	11	5617955	5665628	NA	Yes	rs183322782;rs148179286;rs28576676	6
ENSG00000258659	TRIM34	11	5640994	5665628	NA	No	rs183322782;rs148179286	6
ENSG00000132274	TRIM22	11	5710919	5758319	NA	Yes	rs28576676	6
ENSG00000110148	CCKBR	11	6280966	6293357	NA	Yes	rs113892119:rs28576676	6
ENSG00000170955	PRKCDBP	11	6340176	6341877	NA	Yes	rs28576676	6
ENSG00000108960	MMD	17	53469974	53499353	Yes	No	rs8176751	5
ENSG00000196628	TCF4	18	52889562	53332018	Yes	No	rs687621	5

S5 Data. Severe malaria candidate genes identified by gene-based GWAS analysis using Pascal method

Gene	chromosome	Strand	Cytoban	start	End	Biotype	N of SNPs	P-value
ABLIM2	chr4	-	p16.1	7967036	8160559	protein_coding	37	3.80E-09
ABO	chr9	-	q34.2	136130562	136150630	protein_coding	169	2.67E-11
ADAMTS16	chr5	+	p15.32	5140442	5320412	protein_coding	48	1.61E-07
ARHGAP18	chr6	-	q22.33	129898239	130182692	protein_coding	53	1.00E-08
ATP2B4	chr1	+	q32.1	203595914	203713209	protein_coding	106	4.68E-09
ATXN7L1	chr7	-	q22.3	105245220	105517031	protein_coding	80	3.40E-08
BACE2	chr21	+	q22.2	42539727	42648524	protein_coding	23	3.78E-07
CACNA1A	chr19	-	p13.13	13317255	13617274	protein_coding	41	2.55E-06
CAMK1D	chr10	+	p13	12391582	12871733	protein_coding	41	1.87E-06
CDH13	chr16	+	q23.3	82660398	83830215	protein_coding	245	4.19E-08
CLYBL	chr13	+	q32.3	100258918	100549388	protein_coding	16	5.83E-07
CNOT6	chr5	+	q35.3	179921416	180005353	protein_coding	33	3.23E-07
CNTN4	chr3	+	p26.3	2140549	3099645	protein_coding	114	3.88E-09
CNTN5	chr11	+	q22.1	98891705	100229616	protein_coding	240	1.28E-06
CNTNAP2	chr7	+	q35	145813452	148118088	protein_coding	158	6.10E-08
COL27A1	chr9	+	q32	116918230	117072975	protein_coding	32	2.72E-06
CPXM2	chr10	-	q26.13	125465725	125699779	protein_coding	17	2.69E-06
CRBN	chr3	-	p26.2	3191316	3221401	protein_coding	25	1.23E-07
CSMD1	chr8	-	p23.2	2792874	4852328	protein_coding	499	1.58E-12
CTIF	chr18	+	q21.1	46065426	46389586	protein_coding	22	1.71E-07
DAB1	chr1	-	p32.1	57463578	59012446	protein_coding	136	5.08E-07
DGKB	chr7	-	p21.2	14184673	14942550	protein_coding	52	3.87E-09
DIP2C	chr10	-	p15.3	320129	735608	protein_coding	104	4.54E-07
DLGAP1	chr18	-	p11.31	3496029	4455266	protein_coding	72	2.38E-10
ERICH1-AS1	chr8	+	p23.3	687586	1087777	protein_coding	64	2.04E-06
EVC	chr4	+	p16.2	5712923	5830772	protein_coding	16	7.56E-07
FHIT	chr3	-	p14.2	59735035	61237133	protein_coding	179	1.78E-09
FLT4	chr5	-	q35.3	180028505	180076624	protein_coding	25	9.96E-08
FRMD4A	chr10	-	p13	13685705	14504143	protein_coding	136	1.69E-07
FZD10	chr12	+	q24.33	130647003	130650285	protein_coding	18	1.26E-07
FZD10-AS1	chr12	-	q24.33	130636137	130646768	lncRNA	18	1.26E-07
GAS7	chr17	-	p13.1	9813925	10101868	protein_coding	19	1.63E-06
GLIS3	chr9	-	p24.2	3824127	4300035	protein_coding	25	1.50E-07
GPC5	chr13	+	q31.3	92050934	93519487	protein_coding	93	1.59E-06
HBB	chr11	-	p15.4	5246695	5248301	protein_coding	177	4.92E-10

HBBP1	chr11	-	p15.4	5263184	5264822	transcribed_unprocessed_pseudogene	108	2.00E-08
HBD	chr11	-	p15.4	5254058	5255858	protein_coding	150	8.79E-10
HBE1	chr11	-	p15.4	5289579	5526882	protein_coding	325	9.39E-11
HBG2	chr11	-	p15.4	5269501	5667011	protein_coding	478	1.00E-12
ICA1	chr7	-	p21.3	8152814	8302242	protein_coding	60	4.63E-07
IL2RA	chr10	-	p15.1	6052656	6104333	protein_coding	35	1.31E-06
KAZN	chr1	+	p36.21	14925212	15444544	protein_coding	44	2.15E-06
KCNIP1	chr5	+	q35.1	169780880	170163636	protein_coding	42	2.64E-06
KCNIP4	chr4	-	p15.2	20730238	21950374	protein_coding	157	2.31E-06
KCTD1	chr18	-	q11.2	24034873	24237365	protein_coding	17	3.25E-07
KIF26B	chr1	+	q44	245318286	245866428	protein_coding	151	2.23E-08
KIR3DX1	chr19	+	q13.42	55043908	55055195	transcribed_unprocessed_pseudogene	11	1.53E-06
LINC00260	chr1	+	q32.1	203699704	203700979	lncRNA	99	5.36E-09
LINC00856	chr10	+	q22.3	80008381	80311112	lncRNA	16	1.95E-06
LOC100134259	chr2	+		47055002	47086145	lncRNA	13	4.69E-07
LOC100507584	chr6	-		33857287	33864684	lncRNA	7	1.37E-06
LOC388948	chr2	+		47043806	47049799	lncRNA	12	7.44E-07
LPP	chr3	+	q27.3	187871662	188608460	protein_coding	83	5.61E-08
LRP1B	chr2	-	q22.2	140988995	142889270	protein_coding	266	9.36E-09
MAGI2	chr7	-	q21.11	77646373	79082890	protein_coding	52	1.04E-06
MED22	chr9	-	q34.2	136207754	136214972	protein_coding	42	5.65E-08
MEGF6	chr1	-	p36.32	3404505	3528059	protein_coding	45	6.37E-07
MIR5188	chr12	+	q24.31	125400092	125400205	miRNA	26	1.10E-06
MTHFD1L	chr6	+	q25.1	151186814	151423023	protein_coding	36	4.30E-08
MTUS2	chr13	+	q12.3	29598747	30080084	protein_coding	90	2.56E-06
MYO16	chr13	+	q33.3	109248499	109860355	protein_coding	32	9.38E-07
NALCN	chr13	-	q33.1	101706129	102068843	protein_coding	73	1.19E-06
NAV2	chr11	+	p15.1	19372270	20143147	protein_coding	49	2.84E-06
NKAIN2	chr6	+	q22.31	124125068	125146786	protein_coding	68	6.74E-09
NMT2	chr10	-	p13	15147770	15210695	protein_coding	59	2.57E-06
NPAS3	chr14	+	q13.1	33408458	34273382	protein_coding	56	3.34E-07
NTM	chr11	+	q25	131240370	132206716	protein_coding	105	5.67E-07
OBP2B	chr9	-	q34.2	136080665	136084628	protein_coding	112	1.37E-11
OPCML	chr11	-	q25	132284874	133402403	protein_coding	112	1.46E-07
OR51B2	chr11	-	p15.4	5344527	5345582	protein_coding	69	8.96E-08
OR51B4	chr11	-	p15.4	5322243	5323176	protein_coding	81	4.62E-10

OR51B5	chr11	-	p15.4	5362112	5526882	protein_coding	241	2.26E-07
OR51B6	chr11	+	p15.4	5372737	5373676	protein_coding	83	1.50E-07
OR51E2	chr11	-	p15.4	4701400	4719076	protein_coding	33	7.15E-07
OR51F1	chr11	-	p15.4	4790208	4791147	Pseudogene	119	8.00E-07
OR51F2	chr11	+	p15.4	4842615	4843644	protein_coding	118	1.99E-06
OR51T1	chr11	+	p15.4	4903048	4904113	protein_coding	37	1.90E-06
OR51V1	chr11	-	p15.4	5220964	5221930	protein_coding	189	4.38E-09
OR52B4	chr11	-	p15.4	4388492	4389616	protein_coding	133	1.56E-06
OR52K2	chr11	+	p15.4	4470524	4471591	protein_coding	129	2.61E-07
OR52R1	chr11	-		4824662	4825610		119	1.98E-06
OR56A1	chr11	-	p15.4	6047900	6048971	protein_coding	29	2.34E-06
PARK2	chr6	-		161768589	163148834		230	6.33E-08
PCSK5	chr9	+	q21.13	78505559	78977255	protein_coding	73	2.88E-11
PGAP2	chr11	+	p15.4	3818953	3847601	protein_coding	15	1.18E-06
PITPNC1	chr17	+	q24.2	65373396	65693379	protein_coding	28	2.20E-07
PLCB1	chr20	+	p12.3	8095099	8865547	protein_coding	147	2.36E-06
PRKAG2	chr7	-	q36.1	151253200	151574316	protein_coding	24	8.63E-07
PRKG1	chr10	+	q11.23	52750910	54058110	protein_coding	125	1.25E-06
PTPRD	chr9	-	p23	8314245	10612723	protein_coding	320	4.72E-07
PTPRN2	chr7	-		157331749	158380482	Protein coding	102	4.20E-08
PTPRS	chr19	-	q36.3	5158505	5340814	protein_coding	35	2.16E-07
PTPRT	chr20	-	q13.11	40701391	41818557	protein_coding	65	4.92E-07
RBCK1	chr20	+	p13	388708	411610	protein_coding	26	9.65E-07
RBFOX1	chr16	+		6069131	7763340		219	9.76E-11
RBFOX3	chr17	-	q25.3	77085426	77512230	protein_coding	34	9.50E-07
RBM17	chr10	+	p15.1	6130948	6159422	protein_coding	40	5.80E-07
RIMBP2	chr12	-	q24.33	130880680	131200826	protein_coding	78	3.33E-07
RORA	chr15	-	q22.2	60780482	61521502	protein_coding	63	2.28E-07
RPL7A	chr9	+	-	136215068	136218280	Protein coding	38	5.81E-08
RUNX1	chr21	-	q22.12	36160097	37357047	protein_coding	115	5.70E-07
RYR2	chr1	+	q43	237205701	237997288	protein_coding	115	1.74E-08
RYR3	chr15	+	q13.3	33603176	34158303	protein_coding	93	1.04E-07
SARDH	chr9	-	q34.2	136528683	136605077	protein_coding	8	2.69E-06
SCARB1	chr12	-	q24.31	125262173	125365007	protein_coding	33	9.95E-07
SCGB3A1	chr5	-	q35.3	180017104	180018487	protein_coding	24	1.46E-07
SLC35F3	chr1	+	q42.2	234040678	234460262	protein_coding	73	4.50E-08
SLIT3	chr5	-	q35.1	168088737	168728133	protein_coding	30	1.82E-07
SMYD3	chr1	-	q44	245912641	246670644	protein_coding	142	1.79E-06

SNORA77	chr1	+	q32.1	203698708	203698833	snoRNA	101	5.53E-09
SNORD24	chr9	+	q34.2	136216250	136216325	snoRNA	37	1.07E-07
SNORD36A	chr9	+	q34.2	136217310	136217382	snoRNA	37	1.17E-07
SNORD36B	chr9	+	q34.2	136216948	136217019	snoRNA	36	2.15E-07
SNORD36C	chr9	+	q34.2	136217700	136217767	snoRNA	37	1.17E-07
SORCS2	chr4	+	p16.1	7194373	7744564	protein_coding	111	5.30E-09
SSUH2	chr3	-	p25.3	8661085	8786726	protein_coding	41	3.92E-08
SURF1	chr9	-	q34.2	136218665	136223361	protein_coding	37	1.12E-07
SURF2	chr9	+	q34.2	136223420	136228040	protein_coding	34	1.74E-07
SURF4	chr9	-	q34.2	136228339	136242970	protein_coding	31	8.55E-07
SURF6	chr9	-	q34.2	136197551	136203047	protein_coding	62	6.98E-09
TBC1D20	chr20	-	p13	416123	443187	protein_coding	23	2.81E-06
TCEB1	chr8	-		74857372	74884522		27	2.61E-06
TENM3	chr4	+	q34.3	183065139	183724177	protein_coding	53	6.16E-08
THRB	chr3	-	p24.2	24158644	24536453	protein_coding	65	5.38E-07
THSD7B	chr2	+	q22.1	137523114	138435287	protein_coding	77	7.84E-07
TIAM2	chr6	+	q25.2	155153830	155578857	protein_coding	29	2.56E-07
TMEM132C	chr12	+	q24.32	128751947	129192460	protein_coding	62	2.18E-08
TMEM132D	chr12	-	q24.33	129556270	130388212	protein_coding	77	1.17E-06
TMPRSS6	chr22	-	q12.3	37461478	37505603	protein_coding	16	1.25E-06
TRIM21	chr11	-	p15.4	4406126	4414926	protein_coding	133	1.56E-06
TSHZ2	chr20	+	q13.2	51588945	52111869	protein_coding	13	2.42E-06
TSPAN11	chr12	+	p11.21	31079361	31149537	protein_coding	15	1.03E-06
UBC	chr12	-	q24.31	125396191	125399587	protein_coding	24	1.03E-06
VRK1	chr14	+	q32.2	97263683	97347951	protein_coding	9	2.13E-06
VSTM1	chr19	-	q13.42	54544079	54567207	protein_coding	14	2.39E-06
WBSCR17	chr7	+		70597522	71178586		117	4.46E-07
WWOX	chr16	+	q23.1	78133326	79246564	protein_coding	177	1.42E-06
ZNF385B	chr2	-	q31.3	180306710	180726232	protein_coding	96	2.35E-08
ZNF385D	chr3	-	p24.3	21462489	22414123	protein_coding	64	1.90E-06

S6 Data. Proportion of pathogenic SNPs in candidate malaria resistance genes that recorded higher scores in three endemic countries (Kenya, Malawi and Gambia) compared to the global populations as identified by ANNOVAR

S6 Data Severe malaria resistance candidate genes that contain higher proportion of pathogenic SNPs in three malaria endemic populations (Gambia,Kenya and Malawi) compared to the global populations

Gene	Chromosome	Cyto-band	start	end	Biotype	Khoe.San	Niger.Con	Niger_Cong	Niger_Cong	African_	African.Car	Latin_An	Afro.Asia	Afro.Asiati	Afro.Asiatic	Afro.Asia	European_	European	European	European	USA_India	UK_India	East_Asia	South_As	Gambia	Malawi	Kenya
DBN1	5	q35.3	177456608	177474401	protein_codi	3.534	6.36	9.894	10.954	7.42	6.714	8.481	7.067	2.827	3.18	2.827	3.18	2.12	2.827	2.12	2.12	3.534	5.3	3.18	33.333	33.333	33.333
MMD	17	q22	55392622	55421924	protein_codi	0.993	1.241	1.241	1.737	1.241	1.241	1.737	1.241	1.241	0.744	0.744	0.496	0.496	0.744	0.496	0.496	0.744	0.744	0.496	14.286	14.286	14.286
OR51B2	11	p15.4	5323359	5324297	protein_codi	4.673	5.607	8.411	11.215	7.477	9.346	11.215	10.28	6.542	5.607	1.869	1.869	2.804	1.869	3.738	4.673	3.738	4.673	3.738	28.571	28.571	28.571
OR51B4	11	p15.4	5301014	5301946	protein_codi	5.682	12.5	18.182	21.591	11.364	13.636	15.909	12.5	9.091	5.682	3.409	3.409	2.273	6.818	3.409	2.273	2.273	9.091	3.409	100	100	100
OR51B6	11	p15.4	5351508	5352446	protein_codi	50	52.632	57.895	65.789	55.263	50	57.895	50	47.368	42.105	44.737	47.368	47.368	52.632	50	44.737	42.105	50	47.368	85.714	85.714	85.714
OR51E2	11	p15.4	4680171	4697854	protein_codi	3.058	2.141	3.364	5.81	3.67	3.67	5.505	3.67	2.752	2.446	2.141	3.058	3.058	3.67	2.752	2.141	2.141	4.587	2.446	27.273	27.273	27.273
OR51L1	11	p15.4	4994851	5005536	protein_codi	12.308	16.923	15.385	26.154	20	20	16.923	16.923	13.846	7.692	9.231	10.769	9.231	15.385	10.769	9.231	10.769	13.846	10.769	75	75	75
OR52K2	11	p15.4	4449295	4450361	protein_codi	9.302	9.302	10.853	15.504	10.853	10.853	15.504	11.628	5.426	6.977	5.426	4.651	4.651	6.977	4.651	4.651	5.426	5.426	6.202	100	100	100
OR56A1	11	p15.4	6019336	6034338	protein_codi	1.515	3.03	9.091	21.212	3.03	6.061	7.576	16.667	7.576	12.121	6.061	4.545	4.545	6.061	4.545	4.545	6.061	15.152	4.545	50	50	50
RBCK1	20	p13	407498	432139	protein_codi	2.73	3.97	4.715	6.7	3.226	4.715	4.963	3.474	1.985	1.737	1.985	1.241	1.737	2.978	1.737	1.737	2.233	4.218	2.481	16.667	16.667	16.667
RIF1	2	q23.3	151409883	151508013	protein_codi	2.294	4.493	4.589	6.597	3.059	3.633	4.685	2.294	1.816	1.147	1.338	1.721	1.243	3.92	1.721	1.338	1.825	4.398	2.294	25	25	25
TRIM21	11	p15.4	4384897	4393702	protein_codi	4.255	5.319	4.255	8.865	4.965	4.61	6.383	4.61	2.837	2.837	2.482	2.837	3.191	3.546	3.901	2.837	3.546	4.255	3.901	25	25	25
TRIM22	11	p15.4	5689697	5737089	protein_codi	3.052	4.225	4.695	7.512	4.93	5.164	6.338	3.756	2.582	2.582	2.113	2.817	2.817	5.399	2.817	2.582	2.817	3.286	2.582	23.077	23.077	23.077
TRIM34	11	p15.4	5619764	5644398	protein_codi	1.887	2.695	3.235	5.391	3.235	2.695	4.582	2.965	2.156	1.617	1.078	1.348	1.617	2.965	2.156	2.156	2.426	2.156	2.426	13.333	13.333	13.333
TRIM6-TRIM	11	p15.4	5596109	5612958	protein_codi	1.527	2.29	2.799	4.835	2.799	2.29	3.817	2.545	1.781	1.272	1.018	1.272	1.527	2.545	2.036	2.036	2.29	1.527	2.29	11.765	11.765	11.765
TRIM68	11	p15.4	5596725	5644398	protein_codi	2.791	3.721	4.651	8.837	4.186	3.721	3.256	4.186	3.721	2.326	1.86	1.86	2.326	2.791	2.791	1.395	1.86	3.721	2.791	16.667	16.667	16.667
UBQLNL	11	p15.4	5514393	5518431	protein_codi	11.957	13.043	17.391	21.739	15.217	13.043	18.478	14.13	14.13	11.957	9.783	9.783	9.783	15.217	10.87	10.87	11.957	20.652	7.609	80	80	80
VSTM1	19	q13.42	54040825	54063953	protein_codi	3.941	3.448	4.68	6.404	3.695	4.68	3.695	4.68	3.695	2.217	2.956	3.202	3.448	3.448	3.448	2.709	3.202	3.941	2.956	22.222	22.222	22.222

5. CHAPTER FIVE: CONCLUSIONS AND PERSPECTIVES

5.1. Conclusions

P. falciparum malaria remains one of the leading public health problems worldwide. The global tally of malaria in 2018 was estimated at 228 million cases and 405, 000 deaths worldwide [8]. The global malaria eradication efforts have been implemented in malaria endemic areas using conventional strategies including distribution of long-lasting insecticide treated nets (LLINS), indoor residual insecticide spraying, intermittent treatment for pregnant women in high transmission settings [11]. This led to the significant decline of malaria burden in many parts of the endemic regions [12]. Despite the successes gained, the progress towards global malaria elimination is currently challenged by emergence of drug resistant parasites, insecticide resistant mosquitoes and lack of effective vaccine [12].

To overcome these challenges and maintain the current momentum, there is a pressing need of new intervention tools including effective drugs and vaccines, rapid and accurate diagnostic tools, effective insecticides, improved surveillances and rapid responses among others. A comprehensive understanding of the genetic basis of severe malaria resistance can potentially inform the development of new vaccines and therapeutics. In attempt to identify the natural human genetic factors that confer malaria resistance and determine their corresponding molecular mechanisms, multiple studies have been conducted at least for the last half a century. Earlier studies used conventional genetic study approaches including candidate gene association studies and linkage analysis to identify the genetic variants. In the last decade, several GWASs have been conducted in diverse malaria endemic populations and identified a number of novel variants. However, the conventional GWASs cannot pinpoint the causal variants and their biological functions. Furthermore, the performances of GWASs is relatively weak in African populations in general and in malaria resistance datasets in particular.

To address these challenges, we applied various alternative statistical genetic analytic approaches to the malaria resistance GWAS datasets. We showed that malaria resistance is polygenic trait with h^2_g of ~20% and that the causal variants are overrepresented around protein coding regions of the

genome. The reference panel we prepared as part of our analyses can potentially be used for other LD-based studies in African populations. By applying several gene-based, pathway-level and network-level functional analyses to severe malaria resistance GWAS summary statistics (N=17,000) meta-analyzed across eleven populations in malaria endemic regions in Africa, Asia and Oceania, we systematically identified 57 genes located in the known malaria genomic loci and additional 125 genes across the genome. We showed that the identified genes were significantly enriched in malaria pathogenic pathways including multiple overlapping pathways in erythrocyte-related functions, blood coagulations, ion channels, adhesion molecules, membrane signaling elements and neuronal systems. Furthermore, our population genetic analysis revealed that MAF of the SNPs residing in the identified genes are generally higher in the three malaria endemic populations compared to global populations. Overall, our results suggest that severe malaria resistance trait is attributed to multiple genes that are enriched in pathways linked to severe malaria pathogenesis. This highlights the possibility of harnessing new malaria therapeutics that can simultaneously target multiple malaria protective host molecular pathways. These findings laid the foundations of future experimental studies which can potentially lead to translational medicine such as development of vaccine and new therapeutics.

5.2. *Limitations and the way forward*

We noted that the current severe malaria GWAS datasets are comprised of several populations each with small sample size; making it difficult to perform population specific analyses including GWASs and heritability studies to capture local variations. Moreover, our h^2_g enrichment study was very limited as it requires larger sample sizes to produce reliable estimates at cell and pathways levels. The reference panel we prepared as part of our analyses can be further enriched by including additional datasets such as H3Africa datasets in the future. The h^2_g enrichment studies can potentially provide clues about the functional groups, tissues, cells and pathways in which the causal variants are localized. Thus, powered studies with larger sample size from each study population would provide more resolutions to the current findings.

More powered studies would also enable investigation of epistatic interactions of malaria resistance loci at genome-wide scale and genetic correlation between malaria resistance and susceptibility/resistance to other infectious diseases. Such studies can potentially provide clues to common molecular processes between resistance/susceptibility to infectious diseases and may lead

to translation medicine such as multi-purpose vaccine and therapeutics across common infectious diseases. SM is a complex disease with various clinical manifestations including CM, SMA and others which may arise from distinct pathophysiological processes [37,55]. This implies the existence of sub-phenotype specific variants that influence the disease outcomes. However, sub-phenotype analyses were not presented in the current study owing to the lack of adequate sub-phenotype information in the MalariaGen datasets. We anticipate that the future studies will address these challenges and lead to novel discoveries.

Our genotype-based rare variant association analysis in the current study was limited because, such analyses work best in whole genome dataset. The rare variant and burden analyses enable identifications of rare variants that influence malaria resistance trait unlike the conventional GWASs which focus on common variant associations. Future Whole genome-based studies are needed to better understand the contribution of rare variants in malaria resistance trait. Furthermore, the candidate malaria resistance genes that we identified by our Insilco-functional analyses should be substantiated by experimental studies in the future.

As a concluding remark, besides the GWAS-based studies implemented in our current studies, other -omics approaches including whole genome sequencing, copy number variation, whole exome and proteomic and epigenome studies have huge potentials to unravel the genetic and biological basis of malaria resistance. The application of several single '-omics' study will pave ways to the possibility of multi-stage and multi-dimensional integrative multi-omics study that combines multiple data types from the human host, the parasite, the mosquitoes and the environment. The current biotechnological advances and their continuously decreasing cost together with the availability of advanced analytical techniques, will eventually lead to feasibility of systems biology studies and revolutionize malaria research.

REFERENCES

1. Neghina R, Iacobiciu I, Neghina AM, Marincu I. Malaria, a Journey in Time: In Search of the Lost Myths and Forgotten Stories. Am J Med Sci.2010;340: 492–498. doi:10.1097/MAJ.0b013e3181e7fe6c
2. Carter R, Mendis KN. Evolutionary and historical aspects of the burden of malaria evolutionary and historical aspects of the burden of malaria. Clin Microbiol Rev. 2002;15: 564–594. doi:10.1128/CMR.15.4.564
3. Schlagenhauf P. Malaria: From prehistory to present. Infect Dis Clin North Am. 2004;18: 189–205. doi:10.1016/j.idc.2004.01.002
4. Cox FE. History of the discovery of the malaria parasites and their vectors. Parasit Vect. 2010;3: 5. doi:10.1186/1756-3305-3-5
5. Bruce-Chwatt L. Malaria debated. Nature. 1981;292: 781–782.
6. Majori G. The long road to malaria eradication. Lancet. 2000;31: 354.
7. Piperaki EP. Malaria eradication in the European world : Historical perspective and imminent threats. Intech: 2018:315-334 doi:10.5772/intechopen.76435
8. Wold Health Organization.World malaria report 2018,Geneva, Switherland
9. Egan A, Keusch GT, Breman JG. The intolerable burden of malaria : a new look at the numbers. Am J Trop Med Hyg. 2001;64(1-2 Suppl):iv-vii. doi: 10.4269/ajtmh.2001.64.iv.
10. Gallup JL, Sachs JD. The economic Burden of malaria. Am J Trop Med Hyg. 2001;64: 85–96.
11. Wold Health Organization (WHO). Roll Back Malaria Partnership. The global malaria action plan for a malaria free world. 2019; Geneva, Switherland
12. Dhiman S. Are malaria elimination efforts on right track ? An analysis of gains achieved and challenges ahead. Infect Dis Pov; 2019;8: 1–19.
13. World Health Organization(WHO). World Malaria Day 2017. Geneva, Switherland
14. Patouillard E, Griffin J, Bhatt S, Ghani A, Cibulskis R. Global investment targets for

malaria control and elimination between 2016 and 2030. BMJ Glob Health. 2017;2: 1–11. doi:10.1136/bmjgh-2016-000176

15. Newby G, Bennett A, Larson E, Cotter C, Shretta R, Phillips AA, et al. Review The path to eradication : a progress report on the malaria-eliminating countries.Lancet. 2016; 387:1775–1784.
16. Nkumama IN, Meara WPO, Osier FHA. Changes in malaria epidemiology in Africa and new challenges for elimination. Trends Parasitol. 2017;33: 128–140. doi:10.1016/j.pt.2016.11.006
17. Ranson H, Guessan RN, Lines J, Moiroux N, Nkuni Z, Corbel V. Pyrethroid resistance in African anopheline mosquitoes : what are the implications for malaria control ? Trend Parasitol. 2010; 1–8. doi:10.1016/j.pt.2010.08.004
18. Ashely E, Dhoda M, Fairhurt R, Amaratunga C, Lim P, Suon S, et al. Spread of Artemisinin Resistance in *Plasmodium falciparum* Malaria. N Engl J Med 2014;371:411-423
19. Hemingway J, Shretta R, Wells TNC, Bell D, Djimdé A, Achee N, et al. Tools and strategies for malaria control and elimination : What do we need to achieve a grand convergence in malaria ? PloS Biol.2016; 14:1–14. doi:10.1371/journal.pbio.1002380
20. Achidi EA, Agbenyega T, Allen S, Amodu O, Bojang K, Conway D, et al. A global network for investigating the genomic epidemiology of malaria. Nature. 2008;456: 732–737. doi:10.1038/nature07632
21. Neafsey DE, Volkman SK. Malaria genomics in the era of eradication. Cold Spring Harb Med. 2017;7: 1–12.
22. Bannister LH. Plasmodium. Encycl Life Sci. 2009;13:1-12 doi:10.1002/9780470015902.a0001970.pub2
23. John C. Igweh (2012). Biology of Malaria Parasites, Malaria Parasites, Dr. Omolade Okwa (Ed.), ISBN: 978- 953-51-0326-4, InTech. http://www.intechopen.com/books/malaria-parasites/biology-of- malaria-parasites
24. Beier JC. Malaria Parasite Development. Annu Rev Entomol. 1998; 43:519–543
25. Frevert U. Sneaking in through the back entrance : The biology of malaria liver stages. Tren Parasitol. 2004;20: 417-424 doi:10.1016/j.pt.2004.07.007
26. Vaughan AM, Aly ASI, Kappe SHI. Malaria Parasite Pre-Erythrocytic Stage Infection:

Gliding and Hiding. Cell Host Microb. 2008;4:209–218. doi:10.1016/j.chom.2008.08.010

27. Fujioka H, Aikawa M. Malaria parasite and disease. Chem Immunol. 2002;80: 1–26. doi:10.1159/isbn.978-3-318-00816-6

28. Sturm A. Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. Science. 2006;313: 1287–1290. doi:10.1126/science.1129720

29. Muehlenbachs A, Mutabingwa TK, Fried M, Duffy PE. An unusual presentation of placental malaria : a single persisting nidus of sequestered parasites Hum Path. 2007; 38:520–523. doi:10.1016/j.humpath.2006.09.016

30. Bannister LH, Hopkins JM, Fowler RE, Krishna S, Mitchell GH. A Brief illustrated guide to the ultrastructure of *Plasmodium falciparum* asexual blood stages. Parasitol Today 2000;4758: 427–433.

31. Aikawa M, Miller LH, Johnson J. Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. J Cell Biol. 1978;77: 72–81.

32. Adams JH, Hudson DE, Torii M, Ward GE, Wellems TE, Aikawa M, et al. The duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. Cell. 1990;63: 141–153. doi:10.1016/0092-8674(90)90295-P

33. Chitnis CE. Molecular insights into receptors used by malaria parasites for erythrocyte invasion. Curr Opin Hematol. 2001;8: 85–91. doi:10.1097/00062752-200103000-00005

34. Keeley A, Soldati D. The glideosome : a molecular machine powering motility and host-cell invasion by Apicomplexa. Trend Cell Biol. 2004;14: 528–532. doi:10.1016/j.tcb.2004.08.001

35. Weiss GE, Gilson PR, Taechalertpaisarn T, Tham WH, de Jong NWM, Harvey KL, et al. Revealing the sequence and resulting cellular morphology of receptor-ligand interactions during *Plasmodium falciparum* invasion of erythrocytes. PLoS Pathog. 2015;11: 1–26. doi:10.1371/journal.ppat.1004670

36. Tham WH, Healer J, Cowman AF. Erythrocyte and reticulocyte binding-like proteins of *Plasmodium falciparum.* Trend Parasitol. 2012;28: 23–30. doi:10.1016/j.pt.2011.10.002

37. Miller LH, Baruch DI, Marsh K, Doumbo OK. The pathogenic basis of malaria. Nature. 2002;415: 673–679. doi:10.1038/415673a

38. Cowman AF, Crabb BS. Invasion of red blood cells by malaria parasites. Cell.2006;124: 755–766. doi:10.1016/j.cell.2006.02.006

39. Francis SE, Sullivan DJ, Goldberg DE. Hemoglobin metabolism in the malaria parasite *Plasmodium falciparum*. Annu Rev Microbiol. 1997;51: 97–123.
40. Frank H, Margaret EW, Kenneth EG. Evedence for cyclic development and short-lived maturity in the gametocyte of Plasmodium falciparum. Trans Roy Soc Trop Med Hyg.1971;65: 549–559.
41. Baker DA. Malaria gametocytogenesis. Mol Biochem Parasitol; 2010;172: 57–65. doi:10.1016/j.molbiopara.2010.03.019
42. Sinden RE. Gametogenesis of *Plasmodium falciparum* in vitro:an electron microscopic study. Parasitology. 1982;84: 1–11
43. Billker O, Lindo V, Panico M, Etienne A., Paxton T, Dell A, et al. Identification of xanthurenic acid as the putative inducer of malaria development in the mosquito. Nature.1998;392: 289–292.
44. Bennink S, Kiesow MJ, Pradel G. The development of malaria parasites in the mosquito midgut.Cell Microbiol. 2016;18: 905–918. doi:10.1111/cmi.12604
45. Aly ASI, Vaughan AM, Kappe SHI. Malaria parasite development in the mosquito and infection of the mammalian host. Annu Rev Microbiol. 2009;63: 195–221. doi:10.1146/annurev.micro.091208.073403
46. Bousema T, Drakeley C. Epidemiology and Infectivity of Plasmodium falciparum and Plasmodium vivax Gametocytes in Relation to Malaria Control and Elimination. Clin Microbiol Rev. 2011;24: 377–410. doi:10.1128/CMR.00051-10
47. Kevin M, Dayo F, Catherine W, Isiah M, Maria W, Victoria M, et al. Indicators of life-threatening malaria in african children . N Engl J Med. 1995;332: 1399–1404.
48. Black RE, Morris SS, Bryce J. Where and why are 10 million children dying every year ? Lancet.2003;361: 2226–2234.
49. Wold Health Organization (WHO). Severe malaria. Trop Med Int Heal. 2014;19: 7–131. doi:10.1111/tmi.12313
50. Greenwood B, Marsh K, Snow R. Why do some African children develop Severe Malaria? Parastol Today. 1991;7: 277–281.
51. Gazzinelli RT, Kalantari P, Fitzgerald KA, Golenbock DT. Innate sensing of malaria parasites. Nat Rev Immunol. Nature. 2014;14: 744–757. doi:10.1038/nri3742
52. Moxon CA, Gibbins MP, Mcguinness D, Jr DAM, Marti M. New Insights into Malaria

Pathogenesis. Annu Rev Pathol Mech Dis. 2020;15: 315–343.

53. Taylor TE, Molyneux ME. The pathogenesis of pediatric cerebral malaria : eye exams , autopsies , and neuroimaging. Ann NY Acad Sci.2015;1342: 44–52. doi:10.1111/nyas.12690

54. Molyneux ME. Malarial retinopathy : A newly established diagnostic sign in severe malaria. Am J Trop Med Hyg. 2006;75: 790–797.

55. White NJ. Anaemia and malaria. Malar J. 2018; 17:1–17. doi:10.1186/s12936-018-2509-9

56. Looareesuwan S, Davis TME, Supanaranond W, Desakornl V, Silamut K, Krishna S, et al. Erythrocyte survival in severe falciparum malaria. Act Trop.1991;48: 263–270.

57. Buffet P a, Safeukui I, Deplaine G, Brousse V, Prendki V, Turner GD, et al. The pathogenesis of *Plasmodium falciparum* malaria in humans : insights from splenic physiology. Blood. 2013;117: 381–392. doi:10.1182/blood-2010-04-202911

58. Rowe JA, Claessens A, Corrigan RA, Arman M. Adhesion of *Plasmodium falciparum*-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications. Exp Rev Mol Med. 2009;11: e16. doi:10.1017/S1462399409001082

59. del Portillo HA, Ferrer M, Brugat T, Martin-Jaular L, Langhorne J, Lacerda MVG. The role of the spleen in malaria. Cell Microbiol. 2012;14: 343–355. doi:10.1111/j.1462-5822.2011.01741.x

60. Miller JN. Falciparum malaria-infected erythrocytes specifically bind to cultured human endothelial cells. Science.1981;213: 555–556.

61. Rachane U, Birgitta W, Johan C, Klavs B, Motomi T, Masamichi A, et al. *Plasmodium falciparum*-infected erythrocytes form spontaneous erythrocyte rosettes. J Exp Med. 1989;169: 1835–1840.

62. Pain A, Ferguson DJP, Kai O, Urban BC, Lowe B, Marsh K, et al. Platelet-mediated clumping of *Plasmodium falciparum*-infected erythrocytes is a common adhesive phenotype and is associated with severe malaria. Proc Natl Acad Sci. 2001;98: 1805–1810. doi:10.1073/pnas.98.4.1805

63. Deitsch KW, Wellems TE. Membrane modifications in erythrocytes parasitized by *Plasmodium falciparum.* Mol Biochem Parasitol. 1996;76: 1–10. doi:10.1016/0166-6851(95)02575-8

64. Kraemer SM, Smith JD. A family affair: var genes, PfEMP1 binding, and malaria disease.

Curr Opin Microbiol. 2006;9: 374–380. doi:10.1016/j.mib.2006.06.006

65. Smith JD, Gamain B, Baruch DI, Kyes S. Decoding the language of var genes and Plasmodium falciparum sequestration. Trend Parasitol. 2001;17: 538–545. doi:10.1016/S1471-4922(01)02079-7

66. Kyes SA, Kraemer SM, Smith JD. Antigenic variation in Plasmodium falciparum: Gene organization and regulation of the var multigene family. Eukaryot Cell. 2007;6: 1511–1520. doi:10.1128/EC.00173-07

67. Kraemer SM, Smith JD. Evidence for the importance of genetic structuring to the structural and functional specialization of the *Plasmodium falciparum* var gene family. Mol Microbiol. 2003;50: 1527–1538. doi:10.1046/j.1365-2958.2003.03814.x

68. Jensen ATR, Magistrado P, Sharp S, Joergensen L, Lavstsen T, Chiucchiuini A, et al. plasmodium falciparum associated with severe childhood malaria preferentially expresses PfEMP1 encoded by group A var genes. J Exp Med. 2004;199. doi:10.1084/jem.20040274

69. Avril M, Tripathi AK, Brazier AJ, Andisi C, Janes JH, Soma VL. A restricted subset of var genes mediates adherence of Plasmodium falciparum -infected erythrocytes to brain endothelial cells. PNAS. 2012;109: 1782–1790. doi:10.1073/pnas.1120534109

70. Claessens A, Adams Y, Ghumra A, Lindergard G, Buchan CC, Andisi C, et al. A subset of group A-like var genes encodes the malaria parasite ligands for binding to human brain endothelial cells. PNAS. 2012;109: 1772–1781. doi:10.1073/pnas.1120461109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1120461109

71. Fried M, Duffy PE. Adherence of Plasmodium falciparum to Chondroit i n Sulfate A in the Human Platenta. Science 1996;272: 4–6.

72. Turner GD, Morrison H, Jones M, Davis TM, Looareesuwan S, Buley ID, et al. An immunohistochemical study of the pathology of fatal malaria. Evidence for widespread endothelial activation and a potential role for intercellular adhesion molecule-1 in cerebral sequestration. Am J Pathol. 1994;145: 1057–69.

73. Smith JD, Craig AG, Roberts DJ, Hudson-taylor DE, Peterson DS, Pinches R, et al. Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell. 1995;82: 101–110.

74. Clark IA, Alleva LM, Mills AC, Cowden WB. Pathogenesis of malaria and clinically similar conditions. Clin Microbiol Rev. 2004;17: 509–539. doi:10.1128/CMR.17.3.509-

539.2004

75. Dinarello CA. Immunological and Inflammatory Functions of the Interleukin-1 Family. Ann Rev Immun. 2009. 27:519–550 doi:10.1146/annurev.immunol.021908.132612
76. Naik BRS, Branch OH, Woods AS, Vijaykumar M, Perkins DJ, Nahlen BL, et al. Glycosylphosphatidylinositol Anchors of *Plasmodium falciparum* : Molecular characterization and naturally elicited antibody response that may provide immunity to malaria pathogenesis. J Exp Med. 2000;192: 1563–1575.
77. Artavanis-tsakonas K, Riley EM, Artavanis-tsakonas K, Riley EM. Innate immune response to malaria: Rapid induction of IFN- γ from human NK cells by live *Plasmodium falciparum*-infected erythrocytes. J Immunol. 2002;169: 2956–2963. doi:10.4049/jimmunol.169.6.2956
78. Koch O, Awomoyi A, Usen S, Jallow M, Richardson A, Hull J, et al. IFNGR1 gene promoter polymorphisms and susceptibility to cerebral malaria. J Infect Dis. 2002;185:1684–7 2002; 185:1684–1687.
79. Engwerda CR, Mynott TL, Sawhney S, Souza JB De, Bickle QD, Kaye PM. Locally Up-regulated Lymphotoxin α , not systemic Tumor Necrosis Factor α , is the principal mediator of murine cerebral malaria. J Exp Med 2002;195: 1371–1377. doi:10.1084/jem.20020128
80. Gillrie MR, Lee K, Gowda DC, Davis SP, Monestier M, Cui L, et al. *Plasmodium falciparum* Histones induce endothelial proinflammatory response and barrier dysfunction. Am J Pathol 2012;180: 1028–1039. doi:10.1016/j.ajpath.2011.11.037
81. Moxon AC, Alhamdi AY, Storm BJ, Toh CJ, Ko DJY, Murphy EG, et al. Parasite histones mediate blood-brain barrier disruption in cerebral malaria. Clin med. 2020; 96–97.
82. Pal P, Daniels BP, Oskman A, Diamond MS, Klein RS, Goldberg E. *Plasmodium falciparum* Histidine-Rich Protein II compromises brain endothelial barriers and may promote cerebral malaria pathogenesis. Am Soc Microbiol. 2016;7: 1–11. doi:10.1128/mBio.00617-16.Editor
83. Russell JH, Shigehiko U, Masamichi A, Stephen BA, James HL, Andrew ML, et al. Secretion of a malarial Histidine-rich Protein (Pf HRP II) from *Plasmodium falciparum*-infected Erythrocytes. J Cell Biol. 1986;103: 1269–1277.

84. Fox LL, Taylor TE, Pensulo P, Liomba A, Mpakiza A, Varela A, et al. Histidine-Rich Protein 2 plasma levels predict progression to cerebral malaria in malawian children with *Plasmodium falciparum* infection. Infect Dis Soc Am. 2013;208: 500–503. doi:10.1093/infdis/jit176
85. Gowda NM, Wu X, Gowda DC. The Nucleosome (Histone-DNA Complex) is the TLR9-specific immunostimulatory component of plasmodium falciparum that activates DCs. PLoS ONE. 2011;6. doi:10.1371/journal.pone.0020398
86. Parroche P, Lauw FN, Goutagny N, Latz E, Monks BG, Visintin A, et al. Malaria Hemozoin is immunologically inert but radically enhances innate responses by presenting malaria DNA to Toll-like receptor 9. PNAS. 2007;104: 1–6.
87. Kalantari P, Deoliveira RB, Chan J, Corbett Y, Rathinam V, Stutz A, et al. Dual engagement of the NLRP3 and AIM2 inflammasomes by Plasmodium-derived Hemozoin and DNA during malaria. Cell.2013;6: 196–210. doi:10.1016/j.celrep.2013.12.014
88. Skog J, Würdinger T, Rijn S Van, Meijer DH, Gainche L, Sena-esteves M, et al. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nat Cell Biol. 2008;10: 1470–1476. doi:10.1038/ncb1800
89. Couper KN, Barnes T, Hafalla JCR, Combes V, Ryffel B, Grau GE, et al. Parasite-derived plasma microparticles contribute significantly to malaria infection-induced inflammation through potent macrophage stimulation. PLoS Pathog. 2010;6. doi:10.1371/journal.ppat.1000744
90. Mantel P, Hoang AN, Goldowitz I, Potashnikova D, Hamza B, Vorobjev I, et al. Malaria-infected erythrocyte-derived microvesicles mediate cellular communication within the parasite population and with the host immune system.Cell Microb. 2013; 13:521–534. doi:10.1016/j.chom.2013.04.009
91. Mantel P, Hjelmqvist D, Walch M, Kharoubi-hess S, Nilsson S, Ma S, et al. Infected erythrocyte-derived extracellular vesicles alter vascular function via regulatory Ago2-miRNA complexes complexes in malaria. Nat Commun. 2016;7: 1–15. doi:10.1038/ncomms12727
92. Faille D, Combes V, Mitchell AJ, Fontaine A, Alessi M, Chimini G, et al. Platelet microparticles : a new player in malaria parasite cytoadherence to human brain endothelium. FASEB J. 2009;23: 3449–3458. doi:10.1096/fj.09-135822

93. Wagener FADTG, Volk H, Willis D, Abraham NG, Soares MP. Different faces of the Heme-Heme oxygenase system in inflammation. Pharmacol Rev. 2003;55: 551–571. doi:10.1124/pr.55.3.5.
94. Epiphanio S, Pamplona A, Ferreira A, Chora A, Rodrigues CD, Gregoire IP, et al. Heme oxygenase-1 and carbon monoxide suppress the pathogenesis of experimental cerebral malaria. Nat Med. 2007;13: 703–710. doi:10.1038/nm1586
95. Dudgeon S, Lond FRCP, Clarke C, Lond MD. A contribution to the microscopical Histology of malaria. Lancet. 1916; 153–156.
96. White AV, Lewallen S, Kayira K, Car A, Taylor TE. Correlation of retinal haemorrhages with brain haemorrhages in children dying of cerebral malaria in Malawi. Transl Roy Soc Trop Med Hyg. 2001;95: 618–621.
97. Moxon AC, Heyderman SR, Wassmer CS. Dysregulation of coagulation in cerebral malaria. Mol Biochem Parasitol 2009;166: 99–108. doi:10.1016/j.molbiopara.2009.03.006
98. Hollestelle MJ, Mantey EA, Chakravorty SJ, Craig A, Akoto AO, Donnell JO, et al. von Willebrand factor propeptide in malaria : evidence of acute endothelial cell activation. Br J Haematol. 2006;133: 562–569. doi:10.1111/j.1365-2141.2006.06067.x
99. Kraisin S, Naka I, Patarapotikul J, Nantakomol D, Nuchnoi P. Association of ADAMTS13 polymorphism with cerebral malaria. Malar J. 2011;10: 1–8.
100. Yang Y, Loscalzo J. Regulation of Tissue Factor expression in human microvascular endothelial cells by Nitric Oxide. Circulation. 2000; 2144–2148.
101. Sobolewski P, Gramaglia I, Frangos J, Intaglietta M, Van Der Heyde HC. Nitric oxide bioavailability in malaria. Trend Parasitol. 2005;21: 415–422. doi:10.1016/j.pt.2005.07.002
102. Yeo TW, Lampah DA, Kenangalem E, Tjitra E. Decreased endothelial Nitric Oxide bioavailability , impaired microvascular function , and increased tissue oxygen consumption in children with falciparum malaria. J Infect Dis. 2014;210: 1627–1632. doi:10.1093/infdis/jiu308
103. Yeo TW, Lampah DA, Gitawati R, Tjitra E, Kenangalem E, Piera K, et al. Angiopoietin-2 is associated with decreased endothelial nitric oxide and poor clinical outcome in severe falciparum malaria. PNAS. 2008;105: 17097–17102.
104. Tsiang M, Sciences G, Sadler JE. Regulation of thrombomodulin by Tumour Necrosis

Factor-α : Comparison of transcriptional and post-transcriptional mechanisms Blood. 1991;77: 542–550. doi:10.1182/blood.V77.3.542.bloodjournal773542

105. Michelson AD, Wencel-drake JD, Kestin AS, Barnard MR. Platelet activation results in a redistribution of Glycoprotein IV (CD36). Arterioscler Thromb. 1994;14: 1194–1201.
106. Grau GE, Mackenzie CD, Carr RA, Redard M, Pizzolato G, Allasia C, et al. Platelet accumulation in brain microvessels in fatal pediatric cerebral malaria. J Infect Dis. 2003; 461–466.
107. Krafts K, Hempelmann E. From methylene blue to chloroquine : a brief review of the development of an antimalarial therapy. Parasitol Res.2012; 1–6. doi:10.1007/s00436-012-2886-x
108. Randall M.Rackard. The Origins of Antimalarial-Drug Resistance. N Engl J Med. 2014;37: 397–399. doi:10.1056/NEJMp1404358
109. Talisuna AO, Bloland P, Alessandro UD. History , Dynamics , and Public Health Importance of Malaria Parasite Resistance. Am Soc Microbiol. 2004;17: 235–254. doi:10.1128/CMR.17.1.235
110. Bloland PB. Drug resistance in malaria. WHO. 2001. WHO/CDS/CSR/DRS/2001/4
111. Xin-zhuan SU, H ML. The discovery of artemisinin and the Nobel Prize in Physiology or Medicine. Sci Chine Life Sci.2015;58: 1175–1179. doi:10.1007/s11427-015-4948-7
112. Takala-harrison S, Laufer MK. Antimalarial drug resistance in Africa : key lessons for the future. Ann N Y Acad Sci. 2015;1342: 62–67. doi:10.1111/nyas.12766.Antimalarial
113. Djimde A, Doumbo O, Cortese J, Kayentao K, Doumbo S, Diorte Y, et al. A molecular marker for Chloroquine-Resistant Famciparum Malaria. N Engl J Med. 2001;344: 257–263. doi: 10.1056/NEJM200101253440403.
114. Martin RE, Kirk K. The malaria parasite ’s Chloroquine Resistance Transporter is a Member of the Drug /Metabolite Transporter Superfamily. Mol Biol Evol.2004;21. doi:10.1093/molbev/msh205
115. Mccollum AM, Mueller K, Villegas L, Udhayakumar V, Escalante AA. Common Origin and Fixation of Plasmodium falciparum dhfr and dhps Mutations Associated with Sulfadoxine-Pyrimethamine Resistance in a Low-Transmission Area in South America. Antimicrob Agents Chemother. 2007;51: 2085–2091. doi:10.1128/AAC.01228-06
116. Roper C, Pearce R, Bredenkamp B, Gumede J, Drakeley C, Mosha F, et al. Antifolate

antimalarial resistance in southeast Africa : a population-based analysis. Lancet. 2003;361. doi: 10.1016/S0140-6736(03)12951-0.

117. Talisuna AO, Karema C, Ogutu B, Juma E, Logedi J, Nyandigisi A, et al. Personal View Mitigating the threat of artemisinin resistance in Africa : improvement of drug-resistance surveillance and response systems. Lancet Infect Dis. 2012;12. 888-896. doi: 10.1016/S1473-3099(12)70241-4
118. Das D, Phyo AP, Tarning J, Ph D, Lwin KM, Ariey F, et al. Artemisinin Resistance in Plasmodium falciparum Malaria. N Engl J Med. 2009;361: 455–467.
119. Ariey F, Witkowski B, Amaratunga C, Beghain J, Ma L, Lim P, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature. 2016;505: 50–55. doi:10.1038/nature12876.A
120. Anderson JM, Mao S, Sam B, Sopha C, Chuor CM, Nguon C, et al. Spread of Artemisinin Resistance in Plasmodium falciparum Malaria.N Engl J Med. 2014; doi:10.1056/NEJMoa1314981
121. Kamau E, Campino S, Amenga-etego L, Drury E, Ishengoma D, Johnson K, et al. K13-Propeller Polymorphisms in Plasmodium falciparum Parasites from Sub-Saharan Africa. J Infec Dis. 2015;211: 1352–1355. doi:10.1093/infdis/jiu608
122. Uwimana A, Legrand E, Stokes BH, Ndikumana JM, Warsame M, Umulisa N, et al. Emergence and clonal expansion of in vitro kelch13 R561H mutant parasites in Rwanda. Nat Med. 2020;26. doi:10.1038/s41591-020-1005-2
123. World Health Organization. Malaria vaccine technology roadmap Management of small gastrointestinal stromal tumours. Lancet. 2013;382.
124. Webster D, Hill AVS. Progress with new malaria vaccines. Bulletin of the World Health Organization : the International Journal of Public Health 2003;81: 902–909. https://apps.who.int/iris/handle/10665/72157

125. Duffy PE, Gorres JP. Malaria vaccines since 2000 : progress , priorities , products. Vaccines; 2020; 1–9. doi:10.1038/s41541-020-0196-3
126. Olotu A, Fegan G, Wambua J, Nyangweso G et al. Seven-Year Efficacy of RTS,S/AS01 Malaria Vaccine among Young African Children. N Engl J Med 2016; 374:2519-2529

DOI: 10.1056/NEJMoa1515257

127. Collins KA, Snaith R, Cottingham MG, Gilbert SC, Hill AVS. Enhancing protective immunity to malaria with a highly immunogenic virus-like particle vaccine. Sci Rep; 2017; 1–15. doi:10.1038/srep46621

128. Sissoko M , Healy S , Katile A , Omaswa F , Zaidi I , Gabriel E et al. Safety and efficacy of PfSPZ Vaccine against Plasmodium falciparum via direct venous inoculation in healthy malaria- exposed adults in Mali: a randomised, double-blind phase 1 trial. Lancet Infect Dis. 2019;17: 498–509. doi:10.1016/S1473-3099(17)30104-4.

129. Lyke KE, Ishizuka AS, Berry AA, Chakravarty S, Dezure A, Enama ME, et al. Attenuated PfSPZ Vaccine induces strain-transcending T cells and durable protection against heterologous controlled human malaria infection. Proc Natl Acad Sci. 2017;114: 1–6. doi:10.1073/pnas.1615324114

130. Zheng J, Pan H, Gu Y, Zuo X, Ran N, Yuan Y, et al. Prospects for Malaria Vaccines : Pre-Erythrocytic Stages , Blood Stages , and Transmission-Blocking Stages.BioMed Res Int 2019;2019. https://doi.org/10.1155/2019/9751471

131. Sirima SB, Mordmüller B, Milligan P, Ateba U, Kironde F, Atuguba F, et al. A phase 2b randomized , controlled trial of the efficacy of the GMZ2 malaria vaccine in African children. Vaccine.2016;34: 4536–4542. DOI: 10.1016/j.vaccine.2016.07.041

132. Duffy BPE, Pimenta SP, Kaslow DC. Pgs28 belongs to a family of epidermal growth factor-like antigens that are targets of malaria transmission-blocking antibodies. J Experim Med. 1993;177: 505–510.

133. Kaushal CD, Carter R. Characterization of antigens on mosquito midgut stages of *Plasmodium gallinaceum*. ii, comparison of surface antigens of male and female gametes and zygotes. Mol Biochem Parasitol. 1984;11: 145–156.

134. Jong RM De, Tebeje SK, Tadesse FG, Jore MM, Stone W, Bousema T. Immunity against sexual stage Plasmodium falciparum and Plasmodium vivax parasites. Immunol Rev. 2019;293: 1–26. doi:10.1111/imr.12828.

135. Teo Y, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. Nat Rev Genet. 2013;11: 149–160. doi:10.1038/nrg2731.

136. Howes RE, Battle KE, Mendis KN, Smith DL, Cibulskis RE, Baird JK, et al. Global epidemiology of *Plasmodium vivax*. Am J Trop Med Hyg. 2016;95:15–34.

doi:10.4269/ajtmh.16-0141

137. Miller LH, Mason SJ, Clyde DF, McGuire MH. The resistance factor to *Plasmodium vivax* in blacks: The Duffy-Blood-Group Genotype, FyFy. N Engl J Med. 1976;295: 303–304.
138. Moreno A, Caro-aguilar I, Yazdani SS, Shakri AR, Lapp S, Strobert E, et al. Preclinical assessment of the receptor-binding domain of *Plasmodium vivax* Duffy-binding protein as a vaccine candidate in rhesus macaques. Vaccine.2008;26: 4338–4344. doi:10.1016/j.vaccine.2008.06.010
139. Parikh S, Rosenthal PJ. Human genetics and malaria: relevance for the design of clinical trials. J Infect Dis. 2008;198: 1255–1257. doi:10.1086/592223
140. Weatherall DJ, Clegg JB. Inherited haemoglobin disorders: An increasing global health problem. Bull World Health Organ. 2001;79: 704–712. doi:10.1590/S0042-96862001000800005
141. Hedrick P. Population genetics of malaria resistance in humans. Heredity. 2011;10716: 283–304. doi:10.1038/hdy.2011.16
142. Amaratunga C, Fairhurst RM. α+ Thalassemia antagonizes the malaria-protective effects of sickle-cell trait. EBioMedicine. 2014;1: 8–9. doi:10.1016/j.ebiom.2014.10.020
143. Allison AC. Protection afforded by sickle-cell trait against subtertian malarial infection. Br Med J. 1954;1: 290–294. doi:10.1136/bmj.1.4857.290
144. Angelafago FBJ, Weber RE. Hemoglobin structure and function. Physiology 1998;17.
145. Weatherall DJ. Genetic variation and susceptibility to infection: The red cell and malaria. Br J Haematol. 2008;141: 276–286. doi:10.1111/j.1365-2141.2008.07085.x
146. Karlsson EK, Kwiatkowski DP, Sabeti PC. Natural selection and infectious disease in human populations. Nat Rev Genet. 2014;15: 379–393. doi:10.1038/nrg3734
147. Kwiatkowski DP. How malaria has affected the human genome and what human genetics can teach us about malaria. Am J Hum Genet. 2005;77: 171–192. doi:10.1086/432519
148. Currat M, Trabuchet G, Rees D, Perrin P, Harding RM, Clegg JB, et al. Molecular analysis of the beta-globin gene cluster in the Niokholo Mandenka population reveals a recent origin of the beta(S) Senegal mutation. Am J Hum Genet. 2002;70: 207–223. doi:10.1086/338304
149. Verra F, Mangano VD, Modiano D. Genetics of susceptibility to Plasmodium falciparum:

From classical malaria resistance genes towards genome-wide association studies. Parasite Immunol. 2009;31: 234–253. doi:10.1111/j.1365-3024.2009.01106.x

150. Jallow M, Teo YY, Small KS, Rockett KA, Deloukas P, Clark TG, et al. Genome-wide and fine-resolution association analysis of malaria in West Africa. Nat Genet. 2009;41:657–665 doi:10.1038/ng.388
151. Taylor SM, Cerami C, Fairhurst RM. Hemoglobinopathies: Slicing the gordian knot of *Plasmodium falciparum* malaria pathogenesis. PLoS Pathog. 2013;9. doi:10.1371/journal.ppat.1003327
152. Modiano D, Luoni G, Sirima BS, Simporé J, Verra F, Konaté A, et al. Haemoglobin C protects against clinical Plasmodium falciparum malaria. Nature. 2001;414: 305–308. doi:10.1038/35104556
153. Charache S, Ugoretz RJ, Richard J. Pathogenesis of hemolytic anemia in homozygous hemoglobin C disease. J Clin Invest. 1967;46: 1795–1811.
154. Modiano D, Bancone G, Ciminelli BM, Pompei F, Blot I, Simporé J, et al. Haemoglobin S and haemoglobin C: "Quick but costly" versus "slow but gratis" genetic adaptations to *Plasmodium falciparum* malaria. Hum Mol Genet. 2008;17: 789–799. doi:10.1093/hmg/ddm350
155. Piel FB, Howes RE, Patil AP, Nyangiri OA, Gething PW, Bhatt S, et al. The distribution of haemoglobin C and its prevalence in newborns in Africa. Sci Rep. 2013;3: 1–8. doi:10.1038/srep01671
156. Chotivanich K, Udomsangpetch R, Pattanapanyasat K, Chierakul W, Simpson J, Looareesuwan S, et al. Hemoglobin E: a balanced polymorphism protective against high parasitemias and thus severe *P. falciparum* malaria. Blood. 2002;100: 1172–1176.
157. Ohashi J, Naka I, Patarapotikul J, Hananantachai H, Brittenham G, Looareesuwan S, et al. Extended linkage disequilibrium surrounding the hemoglobin E variant due to malarial selection. Am J Hum Genet. 2004;74: 1198–208. doi:10.1086/421330
158. Thein SL. The molecular basis of beta -Thalassemia. Cold Spring Harb Perspect Med. 2013; 1–24.
159. Sullivan JMO, Preston RJS, Regan NO, Donnell JSO. Emerging roles for hemostatic dysfunction in malaria pathogenesis. Blood. 2016;127: 2281–2289. doi:10.1182/blood-2015-11-636464.

160. Galanello R, Origa R. Beta-thalassemia. Orph J Rare Dis. 2010;5: 1–15.
161. Willcox M, Björkman A, Brohult J. Falciparum malaria and beta-thalassaemia trait in northern Liberia. Ann Trop Med Parasitol. 1983;77: 335–347. doi:10.1080/00034983.1983.11811722
162. Ayi K, Turrini F, Piga A, Arese P. Enhanced phagocytosis of ring-parasitized mutant erythrocytes: A common mechanism that may explain protection against falciparum malaria in sickle trait and beta-thalassemia trait. Blood. 2004;104: 3364–3371. doi:10.1182/blood-2003-11-3820
163. Schuh K, Uldrijan S, Telkamp M, Röthlein N, Neyses L. The plasmamembrane calmodulin-dependent calcium pump: A major regulator of nitric oxide synthase I. J Cell Biol. 2001;155: 201–205. doi:10.1083/jcb.200104131
164. Marsh K, Forster D, Waruiru C, Mwangi I, Winstanley M, Marsh V, et al. Indicators of life-threatening malaria in African children. N Engl J Med. 1995;332: 1399–1404. doi:10.1056/NEJM199505253322102
165. Lamonte G, Philip N, Reardon J, Lacsina JR, Chapman L, Thornburg CD, et al. Translocation of sickle cell erythrocyte microRNAs into *Plasmodium falciparum* inhibits parasite translation and contributes to malaria resistance. Cell Host Microb. 2013;12: 187–199. doi:10.1016/j.chom.2012.06.007.
166. Fairhurst RM, Bess CD, Krause MA. Abnormal PfEMP1/knob display on *Plasmodium falciparum*-infected erythrocytes containing hemoglobin variants: Fresh insights into malaria pathogenesis and protection. Microbes Infect. 2012; 14:851-862 doi:10.1016/j.micinf.2012.05.006
167. Luzzatto L. Sickle cell anaemia and malaria. Mediterr J Hematol Infect Dis. 2012; 4(1):e2012065 doi:10.4084/MJHID.2012.065
168. Cappellini M, Fiorelli G. Glucose-6-Phosphate Dehydrogenase deficiency. Lancet. 2008;371:64-74. doi:10.1016/S0140-6736(08)60073-2
169. Nkhoma ET, Poole C, Vannappagari V, Hall SA, Beutler E. The global prevalence of glucose-6-phosphate dehydrogenase deficiency : A systematic review and meta-analysis. Blood Cell Mol Dis. 2009;42: 267–278. doi:10.1016/j.bcmd.2008.12.005
170. Luzzatto L. G6PD deficiency and malaria selection. Heredity. 2012; 456. doi:10.1038/hdy.2011.90

171. Guindo A, Fairhurst RM, Doumbo OK, Wellems TE, Diallo DA. X-Linked G6PD deficiency protects hemizygous males but not heterozygous females against severe malaria. 2007;4: 516–522. doi:10.1371/journal.pmed.0040066
172. Adachi K, Ghory PK, Schwartz E. A monomeric form of pyruvate kinase in human pyruvate kinase deficiency. Proc NatlAcad Sci. 1977;74: 5–8.
173. Min-Oo G, Fortin A, Tam M, Nantel A, Stevenson MM, Gros P. Pyruvate kinase deficiency in mice protects against malaria. Nature.2003;35: 357–362. doi:10.1038/ng1260
174. Ayi K, Min-oo G, Francisco S, Serghides L, Crockett M. Pyruvate Kinase Deficiency and Malaria. N Engl J Med 2008; 358:1805-1810 .doi:10.1056/NEJMoa072464
175. Yamamoto FI, Mcneill PD, Hakomori SI. Genomic organization of human histo-blood group ABO genes. Glycobiology. 1995;5: 51–58. doi:10.1093/glycob/5.1.51
176. Fry AE, Griffiths MJ, Auburn S, Diakite M, Forton JT, Green A, et al. Common variation in the ABO glycosyltransferase is associated with susceptibility to severe *Plasmodium falciparum* malaria. Hum Mol Genet. 2008;15: 567–576. doi:10.1093/hmg/ddm331
177. Yip S. Sequence variation at the human ABO locus. Ann. Hum. Genet. 2002;6: 1–27
178. Cserti CM, Dzik WH. Review article The ABO blood group system and *Plasmodium falciparum* malaria. Blood. 2017;110: 2250–2259. doi:10.1182/blood-2007-03-077602.
179. Corte A, Benet A, Cooke BM, Barnwell JW, Reeder JC. Ability of *Plasmodium falciparum* to invade Southeast Asian ovalocytes varies between parasite lines. Blood. 2004;104: 2961–2966. doi:10.1182/blood-2004-06-2136.Supported
180. Rubin HL, Zhai SEN, Sahr KE, Liu S. Deletion in erythrocyte band 3 gene in malaria-resistant Southeast Asian ovalocytosis. Proc Natl Acad Sci. 1991;88: 11022–11026.
181. Allen SJ, Alexander N, Peto TE, Weatherall D. Prevention of cerebral malaria in children in Papua New Guinea by SoutheastAsian ovalocytosis. Am J Trop Med Hyg. 1999;60: 1056–1060. doi:10.4269/ajtmh.1999.60.1056
182. Genton B, Al-Yaman F, Mgone C, Alexander N, Paniu M, Alpers M. Ovalocytosis and cerebral malaria. Nature. 1986;19.
183. Jaskiewicz E, Jodłowska M, Kaczmarek R, Zerka A. Erythrocyte glycophorins as receptors for Plasmodium merozoites. Parasit Vectors.2019; 1–11. doi:10.1186/s13071-019-3575-8

184. Pasvol G, Waincoat J., Weatherall DJ. Erythrocytes deficient in glycophorin resist invasion by malarial parasite Plasmodium falciparium. Nature. 1982;297.
185. Booth PB, Mcloughlin K. The Gerbich Blood Group System , especially in Melanesians. Vox Sang. 1972;84: 73–84.
186. Maier A, Reeder J, Patel S, Duraisingh M, Kazura J, Zimmerman P, et al. *Plasmodium falciparum* erythrocyte invasion through glycophorin C and selection for Gerbich negativity in human populations. Nature. 2002; 87–92. doi:10.1038/nm807
187. Yang F, Brune JL, Baldwin WD, Barnett RONR, Bowman BH. Identification and characterization of human haptoglobin. Proc Natl Acad Sci. 1983;80: 5875–5879.
188. Moestrup SK, Mùller HJ. CD163 : A regulated hemoglobin scavenger receptor with a role in the anti-infammatory response. Ann Med;36: 347–354. doi:10.1080/07853890410033171
189. Cox SE, Doherty C, Atkinson SH, Nweneka C V, Fulford AJC, Ghattas H, et al. Haplotype Association between Haptoglobin (Hp2) and Hp promoter SNP (A-61C) may explain previous controversy of Haptoglobin and malaria protection. PloS ONE 2007; 1–8. doi:10.1371/journal.pone.0000362
190. Smith JD, Kyes S, Craig AG, Fagan T, Hudson-taylor D, Miller LH, et al. Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte membrane protein-1 identifies a CD36 binding domain. Mol Biochem Parasitol. 1998;97: 133–148.
191. Silverstein RL, Febbraio M. CD36 , a scavenger receptor involved in immunity , metabolism , angiogenesis , and behavior. Physiology. 2009;72: 1–9.
192. Pain A, Urban BC, Kai O, Casals-pascual C, Shafi J, Marsh K, et al. A non-sense mutation in CD36 gene is associated with protection from severe malaria. Lancet. 2001;357: 1502–1503.
193. Marlin SD, Springer TA. Purified intercellular Adhesion Molecule 1 (CAM-l) is a Ligand for Lymphocyte Antigen 1 (LFA-1). Cell. 1987;51: 813–819.
194. Fernandez-Reyes D, Craig AG, Kyes SA, Peshu N, Snow RW, Berendt AR, et al. A high frequency African coding polymorphism in the N-terminal domain of ICAM-1 predisposing to cerebral malaria in Kenya. Hum Mol Genet. 1997;6: 1357–1360.
195. Luckner D. Association of the ICAM : 1 (KILIFI) mutation with protection against severe malaria in lambarene , Gabon. Am J Trop Med Hyg. 1999;6: 776-779

doi:10.4269/ajtmh.1999.61.776

196. Fry AE, Auburn S, Diakite M, Green A, Richardson A, Wilson J, et al. Variation in the ICAM1 gene is not associated with severe malaria phenotypes. Gen Immun.2008; 9:462–469. doi:10.1038/gene.2008.38
197. Ahearn J, Gearon D. Structure and Function of the Complement Receptors, CRl (CD35) and CR2 (CD21). Adv Immunol. 1989;46: 183–219.
198. Cockburn IA, Mackinnon MJ, Donnell AO, Allen SJ, Moulds JM, Baisor M, et al. A human complement receptor 1 polymorphism that reduces *Plasmodium falciparum* rosetting confers protection against severe malaria. PNAS. 2003;101: 272–277.
199. Thathy V, Moulds JM, Guyah B, Otieno W, Stoute JA. Complement receptor 1 polymorphisms associated with resistance to severe malaria in Kenya. Malar J. 2005;4: 1–7. doi:10.1186/1475-2875-4-54
200. Moulds JM, Kassambara L, Middleton JJ, Baby M, Sagara I, Guindo A, et al. Identification of complement receptor one (CR1) polymorphisms in West Africa. Gen Immun. 2000; 325–329.
201. Hill AVS. Evolution, revolution and heresy in the genetics of infectious disease susceptibility. Phil Trans Roy Soc Biol Sci. 2012;367: 840–849. doi:10.1098/rstb.2011.0275
202. Hill A, Allsopp C, Kwiatkowski DP, Anstey N, McMicheal A, Greenwood MB. Common West African HLA antigens are associated with protection from severe malaria. Nature. 1991;352: 595–600.
203. Chevillard C, Henri S, Stefani F, Parzy D, Dessein A. Two new polymorphisms in the human interferon gamma (IFN- γ) promoter. Eur J Immunogenet. 2002;29: 53–56.
204. Cabantous S, Poudiougou B, Traore A, Keita M, Cisse MB, Doumbo O, et al. Evidence That Interferon-γ plays a protective role during cerebral malaria. J Infect Dis. 2005;192.
205. Forton J, Clark TG, Bancone G, Mangano VD, Luoni G, Rockett KA, et al. Interferon regulatory factor-1 polymorphisms are associated with the control of *Plasmodium falciparum* infection. Genes Immun. 2008;9: 122–129. doi:10.1038/sj.gene.6364456
206. Jorgensen TJ, Ruczinski I, Kessing B, Smith MW, Shugart YY, Alberg AJ. Practice of Epidemiology hypothesis-driven candidate gene association studies : Practical design and analytical considerations. Am J Epidemiol. 2009;170: 986–993. doi:10.1093/aje/kwp242

207. Malaria Genomic Epidemiology Network (MalariaGEN). Reappraisal of known malaria resistance loci in a large multicenter study. Nature. 2014;46: 1197–204. doi:10.1038/ng.3107
208. Teare MD, Barrett JH. Genetic linkage studies.Lancet. 2005; 366: 1036-1044
209. Flori L, Sawadogo S, Esnault C, Fre N, Fumoux F, Rihet P. Linkage of mild malaria to the Major Histocompatibility Complex in families living in Burkina Faso. Hum Mol Genet.2003;12: 375–378. doi:10.1093/hmg/ddg033
210. Timmann C, Evans J, Konig IR, Kleensang A, Ruschendorf F, Julia L. Genome-wide linkage analysis of malaria infection intensity and mild disease. PLoS Genet. 2007;3. doi:10.1371/journal.pgen.0030048
211. Sakuntabhai A, Ndiaye R, Casademont I, Peerapittayamonko C, Rogier C, Tortevoye P et al. Determination and Linkage Mapping of *Plasmodium falciparum* malaria related traits in Senegal. PLoS ONE 3: e2000. doi:10.1371/journal.pone.0002000
212. McGuire W, Hill V. A, Allsopp E. C, Greenwood MB, Kwiatkowski D. Variation in the TNF-α promoter region associated with susceptibility to cerebral malaria. Nature. 1994;371.
213. Brisebarre A, Kumulungui B, Sawadogo S, Atkinson A, Garnier S, Fumoux F, et al. A genome scan for *Plasmodium falciparum* malaria identifies quantitative trait loci on chromosomes 5q31, 6p21.3, 17p12, and 19p13. Malar J. 2014;13: 1–7. doi:10.1186/1475-2875-13-198
214. Garcia A, Marquet S, Bucheton B, Hillaire D, Cot M, Fievet N, et al. Linkage analysis of blood *Plasmodium falciparum* levels : interest of the 5q31-q33 chromosome region. Am. J Trop Med Hyg.1998;58: 705–709.
215. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009;461: 747–753. doi:10.1038/nature08494
216. Reich DE, Lander ES. On the allelic spectrum of human disease.Tren Genet .2001;17: 502–510.
217. Visscher PM, Wray NR, Zhang Q, Sklar P, Mccarthy MI, Brown MA, et al. 10 Years of GWAS Discovery : Biology , Function , and Translation. Am J Hum Genet. 2017;101: 5–22. doi:10.1016/j.ajhg.2017.06.005

218. Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN. Heritability of malaria in Africa. PLoS Med. 2005; doi:10.1371/journal.pmed.0020340
219. Band G, Rockett KA, Spencer CCA, Kwiatkowski DP, Band G, Si Le Q, et al. A novel locus of resistance to severe malaria in a region of ancient balancing selection. Nature. 2015;526: 253–257. doi:10.1038/nature15390
220. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012;489: 443–446. doi:10.1038/nature11334
221. Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, et al. Imputation-based meta-analysis of severe malaria in three African populations. PLoS Genet. 2013; doi:10.1371/journal.pgen.1003509
222. Malaria Genomic Epidemiology Network (MalariaGEN). Insights into malaria susceptibility using genome- wide data on 17,000 individuals from Africa, Asia and Oceania. Nat Commun. 2019;10: 5732. doi:doi.org/10.1038/s41467-019-13480-z |
223. Ravenhall M, Campino S, Sepu N, Nadjm B, Mtove G, Wangai H, et al. Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania.PloS Genet. 2018; 1–15.
224. Parker M, Bull SJ, Vries J De, Agbenyega T, Doumbo OK, Dominic P. Ethical data release in Genome-Wide Association Studies in developing countries. PloS Med.2009;6: 4–7. doi:10.1371/journal.pmed.1000143
225. Timmann C, Thye T, Vens M, Evans J, May J, Ehmen C, et al. Genome-wide association study indicates two novel resistance loci for severe malaria. Nature. 2012;489: 443–446. doi:10.1038/nature11334
226. Loh P, Tucker G, Bulik-sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015; 47,284–290 doi:10.1038/ng.3190
227. Erbe M, Hayes BJ, Matukumalli LK, Goswami S, Bowman PJ, Reich CM, et al. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J Dairy Sci. 2014;97: 6622. doi:10.3168/jds.2014-97-10-6622
228. Zhou X, Stephens M. Genome-wide Efficient Mixed Model Analysis for Association

Studies. Nat Genet. 44: 821–824. doi:10.1038/ng.2310

229. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait analysis. Am J Hum Genet. Am J Hum Genet. 2011;88: 76–82. doi:10.1016/j.ajhg.2010.11.011
230. Finucane HK, Bulik-sullivan B, Gusev A, Trynka G. Partitioning heritability by functional category using GWAS summary statistics.Nat Genet. 2015;47: 1228–12352.
231. Bulik-sullivan BK, Loh P, Finucane HK, Ripke S, Yang J, Working S, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015; 47:291–295 doi:10.1038/ng.3211
232. Watanabe K, Taskesen E. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017; 8: 1–10. doi:10.1038/s41467-017-01261-5
233. Chimusa ER, Mbiyavanga M, Mazandu GK, Mulder NJ. AncGWAS: A post genome-wide association study method for interaction, pathway and ancestry analysis in homogeneous and admixed populations. Bioinformatics. 2015;32: 549–556. doi:10.1093/bioinformatics/btv619
234. Leeuw CA De, Mooij JM, Heskes T, Posthuma D. MAGMA : Generalized Gene-Set Analysis of GWAS Data. PLoS Comput. 2015; 1–19. doi:10.1371/journal.pcbi.1004219
235. Leffler EM, Band G, J Busby GB, Kivinen K, Si Le Q, Clarke GM, et al. Resistance to malaria through structural variation of red blood cell invasion receptors. Science.2017; 356 doi:10.1126/science.aam6393